Smart Energy Feedback in the Home:

The Effect of Disaggregation and Visualisation on Householders' Comprehension of Electricity Data

Melanie Rosemarie Herrmann

UCL Interaction Centre

Department of Computer Science

University College London

Supervisors: Dr. Duncan Brumby, Prof. Tadj Oreszczyn

Examiners: Dr. Alexa Spence (external), Michelle Shipworth (internal)

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at University College London

2018

Declaration

I, Melanie R. Herrmann, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature: ______________________

Date: ______________________

Abstract

Worldwide, households are being equipped with Smart Meters (SM) and other smart Residential Energy Feedback Systems (REFS). One of the aims of these systems is to help householders understand their consumption and to save money on their bill and to become more sustainable. This thesis examines how users make sense of smart electricity feedback, focusing on two aspects of how feedback is given. One aspect is the role of disaggregation, i.e. the provision of information about the energy consumption of individual household appliances in the home as opposed to aggregate feedback on total household consumption. The other aspect is how to visualise residential electricity data. To investigate these aspects, a mixed methods approach was taken: five qualitative interview studies and three lab experiments were conducted to investigate the impact of these two aspects of residential electricity feedback on householders' understanding of how much energy everyday household practices consume. The evidence of the studies in this thesis suggests that interactive visualisations that show information about the energy use of individual appliances in the home are most useful for householders to understand and possibly decrease their consumption.

Acknowledgements

First and foremost, I would like to thank my supervisors Duncan Brumby and Tadj Oreszczyn. I have been lucky to work with two mentors who have complemented each other so well. Thank you for your continuous support and guidance, your honest feedback, and for keeping me on track.

I would also like to thank my other collaborators. This research was supported by EPSRC grant EP/K504610/1 in partnership with EDF Energy UK Research & Development. My industrial collaborators Charles Delalonde, David Ferguson, and Will Selby, for their invitations to discuss my research with the R&D team, for their enthusiasm and insightful thoughts, and for the opportunity to run a study together.

Enrico Costanza, for his mentorship since joining UCLIC. I greatly appreciate the time and effort and the invaluable input, thank you for allowing me to work with you.

Xavier Gilbert and Longmin Cheng, for their contributions and hours spent in the labs.

Paul Marshall and Yvonne Rogers, for examining my registration and upgrade viva. The examinations, and beyond that the numerous chats with Paul, helped me to learn how to defend my work, and they ensured it was moving in a sensible direction.

Alexa Spence and Michelle Shipworth, for the time and effort to examine my viva, and for their thorough critique and encouragement, which made my viva an enriching experience.

Everyone in UCLIC, not only for providing feedback and reading drafts, but above all for the everyday chats, kind words, and friendships. I feel lucky that I have been in this environment for the past years and I will miss you.

Finally, my mum, my family, and my friends, for bearing with me, for their encouragement, support, and patience. After 20-odd years I won’t be a student any longer. And my partner James, for allowing me to turn the kitchen into a writing workshop, for supporting me in figuring out where to go from here, and for doing the dishes in the past weeks.

Impact Statement

This work is motivated by climate change mitigation and the need to reduce global carbon emissions. Residential energy usage makes up 20-30% of global energy consumption. Over recent years, there has been an international rollout of Smart Meters with In-Home Displays designed to try and help people better understand and reduce their consumption.

This thesis investigates householders' interaction with smart energy feedback systems to examine how they make sense of the feedback and to understand how smart technologies can be improved to support householders in reducing energy consumption. This research has produced new knowledge about user needs and design implications for smart energy feedback. This is relevant to academics researching energy feedback and sustainability. The work in this thesis has been presented at relevant international conferences and published in leading peer-reviewed journals.

Outside of academia, this research has directly translated into a field trial with EDF Energy UK, one of the big six energy suppliers in the UK. In a pilot study, energy feedback based on the findings of this work has been tested in the field. This research can inform research and development in the energy sector and policies defining the specifications and the design of future generations of smart technology.

From an economic point of view, better feedback is a competitive service between energy providers. From the societal and environmental perspective, the benefit lies in helping consumers to better understand their energy use and to reduce their consumption.

Publications

Part of the work in this thesis has been peer-reviewed and published. The following overview lists the published/submitted journal articles:

Herrmann, M. R., Brumby, D. P., Cheng, L., Oreszczyn, T., & Gilbert, X. M. (in revision). Visualizing domestic energy data: disaggregate appliances but avoid time series graphs. Submitted to International Journal of Human-Computer Studies.

Herrmann, M. R., Brumby, D. P., Oreszczyn, T., & Gilbert, X. M. (2017). Does data visualization affect users' understanding of electricity consumption? Building Research & Information, 1-13. DOI: 10.1080/09613218.2017.1356164

Herrmann, M. R., Brumby, D. P., & Oreszczyn, T. (2017). Watts your usage? A field study of householders' literacy for residential electricity data. Energy Efficiency, 1-17. DOI: 10.1007/s12053-017-9555-y

Collaborations

Studies 1, 2 and 3 were designed, conducted and analysed by the author of this thesis.

Study 4 was designed and conducted by Enrico Costanza (collaborator at UCL). The unanalysed raw data from this study was kindly shared with the author of this thesis, who then conducted and reported on a novel analysis of these data.

Study 5 was designed, conducted and analysed by the author of this thesis. The experiment was programmed by Xavier M. P. Gilbert (collaborator at Imperial College London). The participant testing for Study 5.3 was carried out by Longmin Cheng (MSc HCI student at UCL).

Study 6 was designed, conducted and analysed by the author of this thesis. This study has been carried out in collaboration with Will Selby (collaborator at EDF Energy UK R&D). EDF also supported this research by helping with the recruitment of the sample and providing the technical sensing infrastructure and web platform for hosting the Residential Energy Feedback System.

Studies 1, 2, 3, 5 and 6 have been supervised by Duncan Brumby and Tadj Oreszczyn.

Research Ethics

The researcher had ethical approval from University College London for the work conducted for this thesis (UCLIC/1415/002/Staff Brumby/Herrmann) and all studies complied with the American Psychological Association's (APA) Ethical Principles of Psychologists and Code of Conduct. In all studies, information sheets and consent forms in line with UCL standards were used. The collected data has been treated as strictly confidential and handled in accordance with the provisions of the Data Protection Act 1998.

Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Motivation

Anthropogenic climate change is one of the biggest challenges of the 21st century according to the United Nations (www.un.org). Residential energy use makes up 25-29% of total energy consumption in the US and UK respectively and this share is forecasted to grow (Mogles et al., 2017). At the same time, industrialised countries are expected to reduce total emissions by 80% below 1990 levels by 2050 to achieve international carbon targets (Cosar-Jorda, Buswell, & Mitchell, 2013). The field of Sustainable Human-Computer Interaction (SHCI) contributes to sustainability by providing tools to track the environmental impact of human activity. Thus, it creates the opportunity to reflect on personal responsibility and inspires behaviour change.

Smart technology may facilitate saving by informing users about their individual energy consumption. Eco-feedback and its effect on behaviour change has been studied for many decades (Darby, 2006). The hope is that eco-feedback from smart new products will be more effective than ever in helping householders to understand and reduce their consumption (DECC, 2013, www.gov.uk). Utility companies for example can replace estimated energy costs by accurately calculated bills. This improvement directly arises from the possibilities created by smart metering.

However, empirical evidence suggests that eco-feedback from smart technologies is not always yielding the expected outcomes. The scientific community has questioned how much smart energy feedback contributes to fighting climate change (Brynjarsdottir et al., 2012; Hargreaves, 2018). It seems that despite near real-time feedback and high temporal resolution of the recorded data, householders do not necessarily make better choices in their everyday energy consumption.

Appliance-wise disaggregation, which gives feedback about the energy consumption of the individual devices in a home, has been proposed as an added value which could determine the success of future generation smart infrastructures (Armel, Gupta, Shrimali, & Albert, 2013). Surprisingly, there is a lack of conclusive evidence that disaggregated eco-feedback leads to energy savings (Kelly & Knottenbelt, 2016). This is surprising as one would think it provides householders with useful feedback to help them make informed decisions.

This thesis investigates the question: Do householders understand the electricity feedback provided by smart metering infrastructures? Data feedback as such has no consequences, if users do not understand the information they receive, and if they don't understand the feedback, then they cannot make informed decisions or change their behaviour for the better, even if they want to (Froehlich, 2011; Mettler-Meibom & Wichmann, 1982). There is a rich record of studies considering the effect of feedback on behaviour change, but there is little research on how users read and process energy data. A lack of understanding might explain why smart meter feedback has been found to be fairly ineffective. In particular, we focus here on trying to understand why attempts to provide disaggregated energy feedback have been so surprisingly ineffective in helping people to better understand and reduce their consumption.

1.2 Research Question

Smart Meters with In-Home Displays (IHDs) are being rolled out in many countries worldwide. It is hoped that the smart feedback helps people understand their energy use better and reduce consumption. This thesis is guided by the overarching research question:

Overarching RQ: Do householders understand smart electricity feedback?

There is a sub-set of relevant research questions to investigate how householders interact with and make sense of smart electricity feedback. These questions are listed below.

RQ1: What is energy literacy?

To investigate householders understanding of smart electricity feedback, we first need to operationalise a way of defining what people know about energy. A person's knowledge about energy is generally referred to as *energy literacy*. A review of the literature will show that energy literacy has been defined very broadly in terms of knowledge, attitudes and behaviours to do with energy generation, distribution and use. The first research question in this thesis is to evaluate, and redefine if necessary, the meaning of energy literacy and to establish an actionable user-friendly definition of energy literacy that can be used with householders.

RQ2: How do householders interact with smart electricity feedback?

Over recent years, Smart Meters with IHDs have been rolled out across the UK and plenty of commercial smart feedback tools have become available. This enables us to investigate whether householders understand smart energy feedback and how they read and make sense of the information provided by smart energy feedback systems in their home. Can householders link feedback on energy consumption to their everyday activities around the home that use energy? What do they want to know about their energy consumption, and does the smart feedback provide the desired information? Is smart energy feedback equipped to deliver any relevant learnings that could inform behaviour change?

RQ3: How does the design of the data visualisation affect how people make sense of domestic energy data?

Data visualisations are used to provide householders with feedback on their domestic electricity consumption. These visualisations can be presented to householders either via an In-Home Display, a web-based platform, or a smartphone app. A common approach to visualise energy data is to show usage over time. However, there are alternative approaches that show energy consumption for individual appliances in the home. How does the choice of graphical visualisation affect people's ability to make sense of domestic electricity consumption? Do people need to be able to identify how much energy was being used during specific periods throughout the day or is it better to give information about total consumption over a longer period of time? Do people need to know how much energy individual appliances are using or is it enough to provide information on the total consumption in the household?

These research questions are ordered in logical succession and the chapters in this thesis reflect the order, consecutively answering the questions.

1.3 Thesis Structure

The author of this thesis will be referred to as the researcher. This thesis consists of six chapters. These include a literature review, three empirical chapters presenting six studies, and a discussion. The structure of the thesis is outlined in more detail below:

Chapter 2 presents a literature review of previous relevant research. It reviews publications in the areas of energy feedback, behaviour change theory, data literacy, energy literacy, graphical literacy, and appliance-wise disaggregation of residential electricity data. The review identifies gaps in the literature. First, an actionable, user-centred definition of energy literacy in householders is needed. Second, despite a lot of research, it is unclear how householders analyse and reflect on energy feedback. Third, there is no conclusive evidence whether disaggregating energy data to the appliance level adds the expected value to energy feedback.

Chapter 3 reports Study 1 which addresses RQ1: What is energy literacy? This chapter presents the findings from three focus groups with energy experts and end-users. Based on the qualitative group interview data, the chapter defines the term energy literacy for the scope of this thesis and conceptualises actionable energy literacy as practical understanding of how energy is used in the home. This includes for example the knowledge where householders are using most energy and the understanding that an electric kettle uses more power than a fridge but that the fridge will use more energy over time because it is constantly on.

Chapter 4 addresses RQ2: How do householders interact with smart electricity feedback? This chapter presents three interview studies with householders. Study 2 is an interview study with householders who have SMs and IHDs. Participants in Study 3 have been provided with a commercial smart energy feedback tool by the researcher. Study 4 combines qualitative interview data with digital diary entries where participants documented energy use behaviours. The results reveal that householders struggle to understand feedback from IHDs and off-the-shelf energy feedback tools. The studies identify factors that either confine or facilitate users' reflective processes and their understanding of the energy feedback.

Chapter 5 addresses RQ3: How does the design of the data visualisation affect how people make sense of domestic energy data? This chapter contains controlled tests in the lab and a field deployment. Study 5 presents a set of three controlled lab experiments with between-subjects design that systematically evaluate the influence of disaggregation and visualisation on users' understanding of how much energy typical household appliances consume. Study 6 is a field study to evaluate the visualisation that was favoured by the lab experiments. The main finding from these studies is that disaggregated data visualisations that emphasise appliance- and activity-centric feedback are most suited for users to reflect on their energy use and to identify where they are using most energy and how they could save energy.

Chapter 6 provides a summary and discussion of the empirical findings across all six studies and it describes how the findings provide answers to the posed research questions. This chapter puts the studies and findings of this thesis into context by comparing them to the literature. This chapter also discusses the limitations of this thesis and provides reflections on future research and applications in the field of smart energy feedback.

Chapter 2 Literature Review

This thesis addresses the user-centred perspective of how eco-feedback can help achieve energy savings in the home. In this chapter, we review the literature on energy feedback. This includes work on relevant behavioural and cognitive theories and research on behaviour change, data literacy, energy literacy and graphical literacy (as energy data is often fed back to householders in a graphical format), as well as research on energy data disaggregation. This chapters identifies gaps in the existing literature that this thesis seeks to address.

2.1 Energy Feedback

In the past, householders have received conventional paper bills via mail (Kempton & Layne, 1994). Today, most customers still receive energy bills, either on paper, or digitally via email or web-account (Neustaedter, Bartram & Mah, 2013). A growing percentage of the population now has SMs installed and receives IHDs from their utility companies (DECC, 2015). Another fraction of the population, mostly tech-savvy early-adopters, equip their homes with commercial smart home products (as Study 2 and 6 will show). Research often uses commercial devices or custom-build prototypes. This section briefly reviews currently available forms of residential energy feedback.

2.1.1 Conventional Energy Feedback

Energy bills are the conventional feedback on gas and electricity consumption in the home. Typically, customers receive bills from the utility company on a monthly or quarterly basis. In 1982, Kempton and Montgomery conducted interviews with householders to investigate energy saving practices in the home. In this study, participants experimented with the influence of their behaviour on energy consumption. For example, they would watch less TV, and then look for differences between the last and the next energy bill. The effect size of their 'experimental' changes was too small, though, to be reflected in changes in their bill and this left them frustrated.

In another interview study a decade later, Kempton and Layne (1994) investigated in detail how householders read their bills and analyse their energy consumption. They found that for most consumers, dealing with their energy bill merely means verifying the due charge and paying it. Only if the charge deviated notably from previous bill, householders would inquire with the rest of the family if any changes in behaviour could have incurred higher cost. Yet, they would not find an explanation because most variation in cost would be due to changes in tariff, weather or a combination of all three (which could not be told apart in the bill). The cost of analysis was high for participants, because detailed information was not provided by the utility company; some of the participants kept records of meter readings and temperature to aid the interpretation of their consumption.

In summary, Kempton and Layne found that *'conclusions consumers can draw from their analytical efforts are restricted by the form in which they receive price and consumption data and their limited analytical capabilities'.* The authors coined the comparison of energy bills to grocery bills in a hypothetical supermarket that would not use price tags for the products on offer, but only provide shoppers with a cumulative bill over all purchases at the end of the month. Under such a billing regime, customers could not possibly find out how to save money. To be able to make informed decisions, they need sufficient information on prices, quantity, and quality of goods. A lack of information and consequent informed decisions comes with severe implications for reducing energy consumption. If householders do not know where they are using energy (and maybe excessively using energy), they cannot take actions to cut their consumption, even if they want to (Mettler-Meibom & Wichmann, 1982).

These findings from the 80s and 90s have been replicated more recently, and it has been confirmed that utility bills remain inadequate for tracking consumption (Neustaedter, Bartram, & Mah, 2013). For instance, bills still only arrive at the end of the month but householders would need instant feedback to enable them to experiment with and learn how changes in their use of appliances around the home affects their overall consumption. Household members who consult their bills often do not know how to read them (i.e. they find energy units such as kilowatts confusing). However, many household members do not even get to see their utility bill and are therefore unaware of the cost associated with their use of electronic appliances in the home (Chetty, Tran, & Grinter, 2008).

Neustaedter, Bartram, and Mah (2013) interviewed householders on their energy bills and asked them to make sense of them and to relate them to their everyday lives. Participants attributed changes in energy consumption to external factors like weather and temperature. They did not tie the changes to activities within the home, even though they consulted their personal family calendars during the interviews to help understand their consumption patterns. Neustaedter et al. concluded that in the 18 years since Kempton and Layne's initial study, little has changed and householders still do not learn much from energy bills. To remedy this, Neustaedter et al. recommend that energy feedback should provide more information about the activities and patterns of consumption.

2.1.2 Smart Energy Feedback

Given the challenges with conventional bills, it is hoped that smart feedback can provide more suitable feedback. There is a range of tools that fall into the category of smart energy feedback. The following section briefly describes both Smart Meters with In-Home Displays provided by utility companies as part of the nationwide rollout, and other commercial Residential Energy Feedback Systems (REFS). The subsequent sections then outline the hopes that are placed in smart feedback to improve householders' energy management, and the challenges with smart feedback.

2.1.2.1 Smart Meters and In-Home Displays

The Department of Energy and Climate Change (DECC, 2013, www.gov.uk) has defined Smart Meters as:

> *'the next generation of gas and electricity meters and they can offer a range of intelligent functions. Domestic customers will be offered an In-Home Display (IHD) linked to their smart meter, enabling them to see what energy they are using and how much it is costing'.*

The Smart Metering Equipment Technical Specifications (SMETS) describe the minimum capabilities for gas and electricity smart metering systems and the IHDs, which utilities must provide along with the smart meter (DECC, 2013). The SMETS state that the *'IHD shall be designed to enable the information displayed on it to be easily accessed and presented in a form that is clear and easy to understand'*. Furthermore, the SMETS require that both real-time usage information as well as historic usage information be displayed. An example of an IHD is depicted in Figure 1. The default screen displays the cost in Pound sterling (£) of energy consumed so far on a given day. This can be changed to display the cost for the week or month. A traffic light colour coded bar at the top indicates whether the consumption is low (green), medium (amber), or high (red) relative to the household's typical use.

Figure 1. The Smart Meter IHD by EDF Energy UK.

2.1.2.2 Smart Residential Energy Feedback Systems

There are numerous commercial REFS available on the market which householders can buy and setup themselves (for example the Loop energy saving kit, which is used in Study 3, see section 4.8.2). Scientific studies on eco-feedback often build their own prototypes (for example FigureEnergy, which is used in Study 4, see section 4.13.2). In this thesis, we not only consider SMs but any REFS that allow for a two-way communication as described by Darby (2010): the household's energy consumption is transferred to the device provider's database (i.e., a utility company or an independent third party), and usage information is fed back to the household by means of an IHD or an Internet-based service that users can access (i.e., via a webpage or smartphone app).

2.1.3 Benefits of Smart Energy Feedback

DECC declared that the rollout of SMs will play an important part in Britain's transition to a low-carbon economy, helping to meet some of the long-term challenges in ensuring an affordable, secure and sustainable energy supply (www.gov.uk). SMs would provide consumers with more accurate information and bring an end to estimated billing. Thanks to IHDs, consumers would be in control, have near real-time information on their energy consumption to help them manage their energy use, avoid waste, save money, and reduce emissions. Indeed, reviews of empirical studies suggest that smart feedback can lead to energy consumption reduction of between 5-15%, with year-on-year reductions of approximately 3% in long- term studies (Darby, 2006). In a nutshell, SMs and IHDs are meant to solve a lot of problems with conventional energy bills. They provide immediate feedback and make energy consumption visible through a ubiquitous device in the home. Other REFS might not come with a IHD unit, but equally, they provide near-instantaneous feedback via web or mobile platforms, accessible to all household members, thus solving the issue that often only one person who pays the bills looks at the feedback (Chetty et al., 2008).

The benefits of energy feedback have been tested in many studies, and different interventions have investigated a myriad of factors that might impact the efficacy of feedback on behaviour change. Amongst others, these factors are timing and frequency of feedback, historic comparisons to a household's past performance, and social comparisons to other households. Further, the effects of goals, rewards and tailored feedback have been tested, as well as the medium through which feedback is given. Several meta-reviews provide a full report on all factors that have been examined (Abrahamse, Steg, Vlek, & Rothengatter, 2005; Darby, 2001, 2006; Ehrhardt-Martinez, Donnelly, & Laitner, 2010; Fischer, 2008; Katzev & Johnson, 1987; Murugesan, Hoda, & Salcic, 2015; Roberts & Baker, 2003). The meta-reviews identified evidence in favour of factors that are beneficial for householders to learn about their energy use and to make reductions, and this review will focus on the most important findings by Fischer (2008).

Fischer (2008) summarises that the most successful feedback combines the following aspects: feedback should be given frequently, over a long time, and it needs to provide an appliance-specific breakdown. This means that information should be given about how much energy individual devices in the home consume, rather than only giving information on how much total energy a household has used. Furthermore, the feedback must be presented in a clear and appealing way. Ideally, it makes use of computerised and interactive tools. A final factor that households appreciate are comparisons, both historical comparisons (how does a household perform now compared to the past) and normative comparisons (how does the household do in comparison to what could be expected given their home's size and the family's size). However, Fischer points out that while householders are very fond of comparisons, the empirical evidence for the effect on behaviour change is less clear.

A number of more recent studies provide further insights on how feedback can be made more meaningful by providing it immediately and by adding actionable, practice-centred information. Murugesan, Hoda, and Salcic (2015) argue that feedback needs to be displayed continuously in near real time. The feedback should provide information on idle time and standby time of appliances and it should provide information about the causes of energy consumption. Katzeff, Wessman, and Colombo (2017) investigated the potential of households' electricity load balancing and found that a positive framing around adapting household practices, such as reorganising or skipping practices, was successful in comparison with most studies focusing on restricting energy use and feeding back complex facts and figures. Mogles et al. (2017) also found significant effects on householders' behaviour, their knowledge about their energy use and their engagement with the feedback by providing more detailed information on energy use: the feedback included suggestions for behaviour change through actionable and personalised messages in accordance with householders' values, thus helping participants to overcome wasteful behaviours that they wanted to change. Another recent publication, looking at energy feedback in the workplace rather than the home, provides similar evidence (Spence et al., 2018), namely, that energy feedback systems are more effective when they are designed for the user and their social needs. Other than setting goals, the feedback tool in this study encouraged discussions between colleagues through a digital pin board. Throughout the study and beyond its duration, environmental concern of the staff increased and energy consumption in the workplace decreased.

2.1.4 Challenges for Smart Energy Feedback

DECC defined the main consumer benefit of the SM rollout as financial savings following the use of the IHD which would help consumers reduce their consumption (DECC, 2009). The scientific literature has challenged the effectiveness of smart energy feedback based on the inconsistent evidence. Meta-reviews found that the effect varies widely from savings of only 1% up to 20% (Abrahamse et al., 2005; Darby, 2006; Fischer, 2008). Studies that suffer from methodological flaws (such as the lack of a control group and a lack of statistical controls for weather and demographics) typically found bigger effects, while methodologically rigorous studies that include statistical controls and control groups found smaller effects and occasionally, monetary feedback even lead to increases in consumption (Delmas, Fischlein, & Asensio, 2013). There has been limited consideration of what the capacity for an average household to save is (i.e., how much they *could* save if they wanted to) (Cosar-Jorda et al., 2013). Some have even gone so far as to argue that the research focus needs to move away from the demand-side and that persuasive technology is not the right approach to achieve sustainability (Brynjarsdottir et al., 2012; Hargreaves, 2018; Knowles, Blair, Coulton, & Lochrie, 2014). The following sections review aspects that limit the effectiveness of energy feedback.

2.1.4.1 Lack of Consideration for the Social Context of Home Energy Use

While pervasive computing has become a ubiquitous reality, the question of how those tools improve users' everyday lives has been neglected (Vanhulst & Lalanne, 2015). Social practice theory approaches behaviour change by considering the social and cultural context in which behaviours occur (Hargreaves, 2011). Pierce, Fan, Lomas, Marcu, and Paulos (2010) write that there is a lack of evidence to explain how feedback affects householders' experience and their specific behaviours and attitudes in the social setting of home energy use:

> *'surprisingly little is known about what specific conservation behaviors do or do not result in such reported savings, how individuals engage or do not engage with feedback, or why conservation does or does not occur in relation to various types of feedback'.*

It remains unclear which behavioural changes account for the energy reductions in published field studies and how energy feedback systems are 'domesticated' into householders' everyday lives. Social and contextual aspects have proven central, though, in understanding householders' energy consumption behaviour. For example, there are non-negotiable practices, meaning that there are behaviours (e.g. using the washing machine to wash laundry) that householders would not compromise for the sake of saving energy. This is easily explained by people's need for comfort, cleanliness, and convenience (Shove, 2003). Shove explains that most resource-consuming practices in the home are inconspicuous habits and routines, not an expression of whether people are actively committed to sustainability or not. When wanting to clean the house or turning on the heating to be warm, householders don't necessarily reflect on the energy they consume. They use energy as means to an end, i.e. to complete necessary household tasks (Entwistle, Rasmussen, Verdezoto, Brewer, & Andersen, 2015). The need to carry out these chores is often more important to people than prioritising a reduction of their energy consumption (or the financial cost of it).

Even though a lot of work has considered socio-demographic factors, there is still too little focus on the human perspective and the complex social interaction within a home. For example, studies often collect data from only one member per household. Yang, Shipworth, and Huebner (2015) quantified statistically that the attitude of householders explained home heating behaviour. Moreover, they found that the attitudes of both partners in the household explained more variance than the attitude of only one. This underlines that studies need to pay more attention to the complex social interactions that are part of the context in which home energy use takes place.

2.1.4.2 Novelty Effects and Lack of Long-Term Engagement

A consistent finding is that behavioural change in the interest of energy saving is transient. Research has shown that indifference and novelty effects are major factors that stop people from regularly interacting with data provided by residential energy displays. Pierce et al. (2010) observed a lack of engagement where users would not even try to test out the device's functions. Common barriers for users' involvement are technical issues, lack of financial incentives and lack of explanations regarding the system (Rego Teixeira, 2014). People consider the devices interesting and exciting in the beginning, but often rather 'gimmicky' and not relevant enough to fully integrate them into their lives (Schlager, 2015). Within a few months, the interaction wears off and any reductions in energy usage quickly evaporate and return to pre-installation levels (Pereira, Quintal, Barreto, & Nunes, 2013).

2.1.4.3 Bounded Rationality

Strengers (2011a, 2014) describes a design approach in smart energy feedback that relates to the idea of the *homo economicus*. The idea is that the *economic man* (or *resource man*) makes purely rational assessments and decisions, being motivated by maximising utility and focusing on economic profit (or savings). Therefore, the assumption is that the *resource man* is empowered by feedback. Smart meters provide the user with consumption information making them aware of their spending, and then the resource man will interpret the eco-feedback and act as *micro-resource manager* and cut consumption in the household to save energy and money. It seems obvious that financial incentives should encourage people to reduce energy consumption. However, sometimes obvious and rational assumptions are not met by actual behaviour. If financial aspects were of highest priority, households would reduce usage and thereby cost on their own initiative. Yet, utility companies struggle to introduce peak time dependent tariffs and acceptance studies on that topic show that people are reluctant to that pricing model although it would enable them to save money (Dütschke, Unterländer, & Wietschel, 2012).

Camerer and Hogarth (1999) have reviewed the effect of financial incentives in 74 experiments. They found that financial incentives improve performance, but they do not eradicate rational violations. In relation to energy spending, it has been pointed out that consumers might be imperfectly informed about cost incurred by energy usage and that there are conflicting motivations other than cost reduction, such as the need for comfort, cleanliness, and convenience (Delmas et al., 2013; Shove, 2003). Another rational, but undesired effect is the so-called rebound effect: improvements in energy efficiency make energy cheaper, and therefore encourage increased consumption (Sorrell, Dimitropoulos, & Sommerville, 2009). If the major motivation for users was only to save energy and money, they would not increase their consumption despite energy being cheap, but the utility of using energy (e.g. being more comfortable/warmer at home) may lead householders to use more energy the cheaper it is. As mentioned earlier, the social context matters, and people consider trade-offs between cost and comfort and convenience (Shove, 2003).

2.1.4.4 Lack of Understanding Resulting from Poor User-Centred Design

Smart home technologies face a design challenge to better fit into the user's life with its practices, dynamics and routines (Rasmussen, 2016). It has been established that energy feedback is often not well designed for everyday life and therefore fails to reach its full potential (Strengers, 2011a). Poor design is often explained by erroneous assumptions in the design process (Busby & Chung, 2003; Osman, 2011). Such erroneous misconceptions are, for instance, that the user has sufficient knowledge about the system, that the user will actively monitor the system and detect changes, or that the user interprets the information correctly. However, householders do not normally monitor their energy consumption (Chetty et al, 2008), analysing energy-related information may be difficult for them (due to low energy literacy which will be discussed in 2.4), and they might struggle to understand and operate smart devices (Yang, Newman & Forlizzi, 2014).

Research has shown that householders' interaction with energy feedback is insufficient, and that there are difficulties in understanding both how to operate smart devices and in understanding the energy data presented. For example, Yang, Newman, and Forlizzi (2014) interviewed households who were using the Nest Learning Thermostat, a programmable but also self-learning, Wi-Fi enabled smart thermostat. Participants hardly interacted with the device and they did not invest the effort to work out the meaning of menu functions that were unclear to them. To assess and potentially adjust the Nest's schedules, participants would initially look at their energy history which shows how the Nest has been operating on a daily or weekly basis. But over time, participants lost interest and did not look at the history any more. Participants' comments revealed that they did not fully comprehend the Nest and many did not even realise that there were problems with their Nest's settings before these were brought to their attention by the researchers who were running the study.

Hargreaves, Nye, and Burgess (2010) and Wallenborn, Orsini, and Vanhaverbeke (2011) have investigated how people appropriate electricity monitors and what they learn when using them. They, too, found that current electricity displays are often poorly designed for the user. For instance, they represent data mostly in kilowatt-hours or monetary cost, whereas aesthetic graphic representations would be more useful for householders. Goulden, Bedwell, Rennick-Egglestone, Rodden, and Spence (2014) emphasise that the design of smart grids must not only focus on the technology but recognise the user whose engagement with the energy feedback is crucial.

Addressing design shortcomings, there have been user-centred attempts to design feedback that helps householders reflect on their data. Costanza, Ramchurn, and Jennings (2012) have deployed an interactive visualisation prototype, FigureEnergy. The web-based feedback of FigureEnergy would show energy consumption over the day in the form of a line graph, with time of day on the x-axis and power on the y-axis. Participants could annotate the graph to create a digital diary. They could select segments of the graph and label them to keep a record of what they were doing at the time. For example, if they were cleaning the house in the afternoon and the graph showed a peak, they could select that peak with the cursor and add a textual label (e.g., 'vacuum cleaning'). Based on these annotations, the system would calculate the energy consumption of different appliances. The energy consumed by every appliance was then displayed in a summary data visualisation. The summary visualisation used a tree map with icons for each appliance that varied in size dependent on the energy consumed by that appliance. Costanza et al. evaluated how people used the FigureEnergy system, finding that it prompted reflection about domestic consumption. However, a limitation of the study was that participants' annotations might have been very error-prone and their accuracy was not controlled for. The prototype evaluation showed that participants were very dedicated to annotating the graph, but they did not make much use of the summary data visualisation.

2.1.5 Conclusion

Smart technology has a lot of potential to advance energy feedback by providing users with richer information about their consumption. However, there are factors that are limiting its success. One of them is that current feedback is not designed to sufficiently map to the social context of energy use. The second is that we do not yet understand well enough how householders are using smart energy feedback and to what extent they understand it and learn from it. A lack of engagement may cause users to not fully understand the feedback, but vice versa it is also possible that a lack of understanding and meaningfulness of the feedback has a negative effect on engagement. If users do not understand the information, they are less likely to look at it again and they cannot change even if they want to.

2.2 Behaviour Change

Behaviour change includes a process from being aware of a behaviour that a person wants to change to the point in which they enact changes to their behaviour. This thesis builds on a behaviour change model from the HCI domain, which was designed specifically to explain behaviour change in the context of ubiquitous computing. It includes the cognitive component of understanding data, or information, presented by a personal informatics systems. The purpose of this review is to identify steps in the process from feedback to behaviour change that are difficult. This may help to better understand current shortcomings of smart energy feedback and it may provide insights on how to improve feedback.

2.2.1 Personal Informatics and the Role of Reflection

The term *personal informatics* refers to systems that help people collect and reflect on personal information (Li, Dey, & Forlizzi, 2010). Typical topics of personal informatics relate to fitness, health and wellbeing (e.g. Bird, Fozzati, Harrison, & Marshall, 2013; Consolvo et al., 2008; Harrison, Marshall, Bianchi-Berthouze, & Bird, 2015; Kay et al., 2012; Lin, Mamykina, Lindtner, Delajoux, & Strub, 2006). While energy feedback might not typically be addressed by personal informatics research, it shares the same characteristics of sensing and collecting personal information for users to reflect on.

Froehlich (2011), who researched the effect of eco-feedback on behaviour change, describes a circuit between sensing, feedback and the user:

> *'A sensing system senses human behavior, which is input into the feedback system and visualised in order to engage, inform and potentially change behavior. Changes in behavior (or the lack thereof) are then sensed by the underlying sensing system and the loop continues'.*

Froehlich's description refers to sensing and feedback systems, such as personal informatics systems or energy feedback systems and is therefore more suitable in our context than broad behaviour change models. Cox, Bird, and Fleck (2013) describe how personal informatics systems can cause 'digital epiphanies' – insights that are gained by using personal informatics tool and that can induce attitudinal and behavioural change: '*Personal informatics systems measure and display information about personal behaviours and can facilitate reflection and increase self-knowledge*'.

Li, Dey, and Forlizzi (2010) present a model of personal informatics systems, containing the five stages Preparation, Collection, Integration, Reflection, and Action. In the Integration stage, the information must be prepared and processed for the user to reflect on it. In the Reflection phase, users must explore and interact with the information to reflect on it. Li, Dey, and Forlizzi also describe problems that can occur in all stages and create barriers, thus preventing the user from transitioning to the later stages. In the Reflection stage, people might have difficulties in exploring and understanding the presented information. A barrier in the Action stage would be if a person did not know how to put the learned knowledge into action. Li et al. suggest that stages can be system-driven or user-driven. When the system is superior, there is a risk of inaccurate automated analysis and the user feels a perceived loss of control. On the other hand, if the user is left in charge, the burden of analysing complex data is left to him or her.

Epstein, Ping, Fogarty, and Munson (2015) expand the Stage-Based Model of Personal Informatics and propose the Lived Informatics Model of Personal Informatics. They point out that previous approaches focus too much on behaviour change goals and so do not adequately capture the nature of tracking in real life. People might have very different goals for using personal informatics tools. For instance, some people track out of curiosity, without wishing to make changes. Still, reflection and comprehension are central. Furthermore, the model addresses the selection of a tool, the interaction with the tool, and its role in information process.

Both theoretical models of personal informatics agree on the central role of reflection. Research on persuasion-based interventions for sustainable behaviour change has emphasised the importance of reflection in practice (Mamykina, Mynatt, Davidson, & Greenblatt, 2008; Prost, Mattheiss, & Tscheligi, 2015; Purpura, Schwanda, Williams, Stubler, & Sengers, 2011). Ploderer, Reitberger, Oinas-Kukkonen, and Gemert-Pijnen (2014) describe how reflecting about aspects of one's life ideally provides insights, which in turn leads people to reconsider and possibly change behaviours. They found that interactivity with feedback, exploration of the data, and learning were key. Schwartz et al. (2015) have established the idea that it matters what people do with technology, as opposed to what technology does to people. Therefore, how people respond to and reflect on their data is crucial because it determines whether they think for themselves and that enables them to decide if they want to improve their behaviour.

It is important that HCI systems enable engagement and reflection about everyday aspects of people's lives. The most important type of reflection, according to Fleck and Fitzpatrick (2010), is transformative reflection which fosters a change in understanding or practice based on the acquisition of new perspectives. Such new perspectives can be acquired through reflection-in-action and reflection-on-action (Schön, 1987). Reflection-in-action means to reflect at the time of doing, reflection-on-action means reflecting on previous activities. Ubiquitous technology has the potential to facilitate both types. In the case of energy feedback, the user could look at it in the very moment when they are carrying out a certain practice, or at the end of the day or week to look at their consumption over time. Reflection-on-action allows for more extensive interaction, and experimentation with data and smart systems can go further than merely providing data on power use at different times of the day, for example, smart systems may wish to point out patterns in the data and to establish connections to social practices in the home.

Despite the popularity and success of personal informatics, some concerns have been expressed in the literature. For example, Greis, Henze, and Schmidt (2015) draw attention to the fact that the systems are sometimes limited to measurement and data display, and they do not offer enough support to implement changes. Froehlich, Kay, Larsen, and Thomaz (2014) enumerate a list of problems. Among these are concerns that certain data collection approaches can compromise the data and that user interface design can provide misguided feedback and negatively influence user experience. The same applies to home energy feedback. Engaging with energy data bears the potential of gaining insights by reflecting on it, but the way energy data is collected, processed and presented affects the feedback's success.

2.2.2 Conclusion

Theories that address human interaction with information based on big data are relatively young. These models consider how users reflect on and integrate the provided information into their lives. Research on persuasive technology agrees on the central role of reflection to generate transformative insights. It is important for energy feedback research to understand how users reflect on energy data and which barriers can arise that prevent them from understanding it. The following sections review users' ability to engage with and understand data in general (data literacy), energy data specifically (energy literacy), and graphically presented data (graphical literacy).

2.3 Data Literacy

Literacy is the ability to read and write. The term literacy is being used increasingly in conjunction with concepts other than linguistics, for example, both in everyday life and in the scientific literature, terms like *computer literacy* have emerged. Literacy in different domains is referred to as a general skill that allows a person to acquire new knowledge (Boy, Rensink, Bertini, & Fekete, 2014). This review focuses on literacy required in the context of home energy feedback, which is data literacy, energy literacy, and graphical literacy. The review begins with the most general one, namely data literacy.

2.3.1 Definition

Data literacy is the ability to derive information from data, i.e. it describes a person's ability to use data as part of everyday life to support thinking and reasoning to solve problems (Wolff, Gooch, Cavero Montaner, Rashid, & Kortuem, 2016). According to Wolff et al., this includes the ability to ask and answer real-world questions from large and small data sets through an inquiry process, which requires both practical and creative thinking. Ackoff (1989) has described the hierarchical relationship of data, information, knowledge and wisdom as a pyramid. Data sits at the bottom of the pyramid and forms the basis for the other three. If analysed, data becomes meaningful information. This information may generate knowledge. Ultimately, one reaches the top of the pyramid, which is wisdom.

2.3.2 Benefits

It is assumed that people will change their behaviour for the better if they receive information on how they can improve. This is based on the information-deficit model which assumes that people lack knowledge and engagement, but this barrier can be overcome when information is provided (Bager, 2014; Dickson, 2005). Hungerford and Volk (1990) write that the aim of education is to shape people's behaviour, and knowledge is central in forming human action. In the context of this thesis, this means that smart energy feedback is not necessarily sufficient, but without it, it's impossible to learn (Darby, 2006). Given that feedback is provided, the other question is whether people comprehend it, because if they do not understand the feedback and do not learn from it, they are unable to change even if they want to (Mettler-Meibom & Wichmann, 1982). There is potential in smart metering to provide rich data and personalised information to catalyse behavioural change, but only if the data presentation is easy to understand and act upon (Krishnamurti, Davis, Wong-Parodi, Wang, & Canfield, 2013).

2.3.3 Challenges

The challenge is for people to understand the data and information they receive and to act on it:

> *'feedback is only information, that is, data and as such has no necessary consequences at all. Like any fact, its effect on action depends on how it is appraised and what decisions are subsequently made with respect to it' (Latham & Locke, 1991).*

Residential energy feedback systems face the challenge of communicating information based on very rich data (Maréchal & Holzemer, 2015). Numerous studies have considered the effects of eco-feedback on behaviour change, but few explicitly investigate data comprehension and the cognitive process of making sense of the data (Yun et al., 2010).

Zhao, Froehlich, and Landay (2010) write that eco-feedback technology is based on a two-part hypothesis (this is equivalent to the information-deficit model mentioned above). First, that users are not aware and knowledgeable enough to understand their behaviour's consequences. Second, that this awareness and knowledge gap can be overcome with the help of feedback through computerized means (e.g., mobile apps, ambient displays, or web-based visualisations). While technological determinism assumes that technologies will shape people's behaviour in predictable ways (Winner, 1980), Froehlich (2011) clarifies that merely feeding back behavioural data does not guarantee any change, and emphasises the importance of the design of eco-feedback.

2.3.4 Conclusion

Data is the foundation for retrieving information and gaining knowledge. This process includes sophisticated computational analysis for big data sets (such as smart energy data) and it requires a level of literacy from the user. The average householder may not be familiar with handling complex data and this may restrict their ability to understand it. We next review people's ability to deal with energy information specifically (energy literacy).

2.4 Energy Literacy

Studies on energy feedback talk about increasing energy literacy, often without explicitly defining energy literacy. The implicit assumption is that householders become more energy literate when receiving information about their consumption, and in turn will improve the way they consume energy. This section first reviews definitions and measures of energy literacy. It then discusses energy literacy in the context of residential energy feedback.

2.4.1 Definitions

DeWaters, Powers, and Graham (2007) describe energy literacy as energy-related knowledge and they define an energy literate person as someone who has a basic understanding of energy concepts and a sound knowledge base. They further propose that energy literacy is a broad term and encompasses knowledge, citizenship understanding, skills, sensitivity, attitudes, and intentions, involvement, and action. In publications a few years later, DeWaters and Powers (2011) refine this by arguing that energy literacy '*will empower people to make appropriate energy-related choices and embrace changes in the way we harness and consume energy*' and that it includes affective and behavioural aspects (DeWaters, Qaqish, Graham, & Powers, 2013). Most publications that mention energy literacy refer to DeWaters and Powers and adopt their definition. However, there are nuances that differ from the original definition.

Mogles et al. (2017) also build on DeWaters and Powers' earlier work for a definition of energy literacy, arguing that it implies an in-depth understanding of energy consumption. However, Mogles et al. raise the issue that DeWaters' work conflates knowledge about energy with the motivation to reduce it by combining cognitive (knowledge, in-depth understanding), affective (attitudes, values) and behavioural (social practices) components together under the term energy literacy (Mogles et al., 2017, p.441). Mogles et al. argue that these three aspects can be distinguished within a more complex concept of energy literacy.

Schwartz, Denef, Stevens, Ramirez, and Wulf (2013) base their definition of energy literacy on the data collection of a three-year long living lab study in which seven households received smart energy feedback. They write that the data collected in this study revealed how householders appropriated the energy feedback and as part of this process continuous learning occurred. This learning led to an increase in knowledge and competence around participants' energy use. For example, participants became familiar with the wattage of different appliances and knew how high their consumption would be at certain times of a typical day. The authors refer to this increase in understanding everyday domestic electricity consumption as energy literacy.

The definitions and how they are derived vary between authors. While DeWaters offers the richest descriptions, they lack a detailed explanation of why attitudes and behaviours are part of the literacy definition. Mogles et al.'s (2017) criticism of this blend of concepts is justified, given that knowledge, attitude, and behaviour are certainly interrelated but separate concepts. Knowledge is a factor in forming an attitude, and attitude is a factor determining behaviour, but neither knowledge nor attitude are the only predictors of behaviour (Ajzen & Fishbein, 1977). Nonetheless, energy literacy ultimately aims at making informed decisions about energy use, and thus attitude and behaviour are important. These go beyond the knowledge-centric understanding that Schwartz et al. (2013) observed in their participants. At this stage, we are not aware of a detailed discussion of this mismatch in definitions in the literature, and therefore this thesis will define energy literacy for the scope of this thesis (Study 1).

2.4.2 Measures

2.4.2.1 Validated Questionnaires

DeWaters, Qaqish, Graham, and Powers (2013) have designed an energy literacy questionnaire for middle and high school students. This questionnaire has been developed as a diagnostic instrument following psychometric principles and its validity has been evaluated statistically, achieving good validity indices (Cronbach's α up to .83). The questionnaire comprises three scales: a cognitive, an affective, and a behavioural scale. The cognitive scale has 50 items, the affective scale has 17 items, and the behavioural scale has 10 items. An example item from the cognitive scale is the following multiple-choice question (the appendix of the publication does not list all multiple-choice options but only the correct answer):

> *The amount of electrical energy (electricity) we use is measured in units called: kilowatt-hours (kWh).*

An item from the affective scale is the following (to be scored on a 5-point Likert-scale with options including 'strongly agree', 'agree', 'neither agree nor disagree', 'disagree', 'strongly disagree'):

> *Saving energy is important.*

An item from the behavioural scale is the following (to be scored on a 5-point Likert-scale with options including 'always or almost always', 'quite frequently', 'sometimes', 'not very often', 'hardly ever or never'):

> *When I leave a room, I turn off the lights.*

Brewer (2013) researched energy literacy in university students and adopted the energy literacy questionnaire developed by DeWaters et al. (2013). Brewer's adapted questionnaire includes 18 items on energy attitudes, 17 items on energy behaviours, and 13 items on energy knowledge. The response format for the scales remained the same, i.e. multiple choice for the knowledge items and Likert-scales for the attitude and behaviour items.

2.4.2.2 Non-Validated Questionnaires

Studies investigating energy literacy in householders have deployed their own set of measures to assess how energy literate their participants were. Brounen, Kok, and Quigley (2013) have measured the energy literacy of residential households by asking them 1) how much they pay for their monthly gas and electricity bill, 2) the temperature that they usually set their thermostat, and 3) what type of thermal insulation they have added to their current home.

Mogles et al. (2017) investigate the influence of smart energy feedback on householders' energy consumption and energy literacy. To measure energy literacy, they use a questionnaire consisting of six questions that ask participants 1) how much they feel they know about energy, 2) which sources of information have contributed most to their understanding of energy use (such as school education, family members, or the Internet), 3) what renewable energy sources are, 4) which source most of the UK's renewable energy comes from, 5) which action would save most energy in the UK (e.g. turning the lights of versus turning the heating down or cycling instead of driving), and 6) which lighting uses the least amount of energy (choosing from a list of different types of lights).

2.4.2.3 Quizzes, Ranking Tasks, Card Sorting and Sketching

Other measures of energy literacy include quizzes, ranking tasks, card sorting technique, and sketching. In previous studies, participants were asked to rank a list of appliances according to how much power the appliances use or to compare two appliances at a time and indicate which one consumes more (Anderson & White, 2009; Yun et al., 2010, ENLITEN game by http://www.cs.bath.ac.uk/enliten/). Gabe-Thomas, Walker, Verplanken, and Shaddick (2016) have investigated householders' mental models of domestic energy consumption by using a card-sorting technique, in which participants were asked to group appliances that they felt belonged together. Gabe-Thomas et al. found that the clusters that emerged across participants through this card sorting activity were based on activities and locations in the home, but not on energy consumption. Chisik (2011) investigated people's ideas about electricity, consumption, and knowledge about which appliances consume a lot of energy. He asked participants to sketch their mental images of electricity and the electrical infrastructure in their home (results indicated that mental models of electricity are not clear and that people use heuristics such as the size of the device or its duration of use to estimate consumption rates).

2.4.2.4 Critique of Measures

DeWaters has established himself as expert in energy literacy and he is the only one who provides a carefully constructed and validated questionnaire. However, as mentioned previously with regards to energy literacy definitions, it remains to be discussed whether attitudes and behaviours should be part of an energy literacy questionnaire. More pressingly, this thesis needs to identify measures of energy literacy in householders, and many of the items from the validated questionnaires are not useful because they are tailored to a school context. While the other reviewed measures of energy literacy have not undergone the same thorough validation, Yun et al. (2010) report convergent validity of their measure, participants' self-assessment, and participants' performance in the study. All measures seem to accurately capture the construct and multiple authors found ranking tasks and quizzes suitable to test participants' energy literacy. The measures have good face validity and are well suited for the context of the research for this thesis.

2.4.3 Energy Literacy in the Home

DeWaters, Qaqish, Graham, and Powers' (2013) definition and inventory have been developed and evaluated in the context of a school curriculum, finding that secondary students have poor energy literacy. Brewer (2013), who builds on this work, finds the same applies to university students. How does the concept of energy literacy apply in the context of energy feedback in the home?

Kempton and Montgomery (1982) found in an ethnographic study that householders made ineffective choices to save energy in the home. These were due to errors in judging how much impact conservation actions would have. For example, participants thought lighting was one of the major uses of energy in the home, even though lighting constitutes a marginal fraction of the total energy used in most homes. Kempton and Montgomery explain this finding by the salience and visibility of lighting, meaning that people use availability heuristics and name things that come to mind very easily (Tversky & Kahneman, 1973, 1974). Further, running time of appliances was equalled to high consumption, regardless of the power load (which might also explain the overestimation of lights, because they are often on for long durations but they are not high-power devices). Another reoccurring heuristic was based on the amount of (human) labour that an appliance replaced, for example, the dishwasher was commonly named as a high-energy consumer. None of the participants in this study mentioned the physics of energy in the interviews and neither did they talk about their residential consumption in terms of kilowatt-hours (which could be expected because this is the commercial units used on the energy bill). Even participants who showed technical understanding, preferred monetary units over energy units and fell prey to the same misjudgements as less knowledgably householders.

Recent research has shown that people have not become more energy literate over the past decades, and they have vague ideas at best of how much energy they are consuming for everyday actions in the home (Bager, 2014; Chisik, 2011; Darby, 2006; Froehlich et al., 2011; Rego Teixeira, 2014). Attari, DeKay, Davidson, and Bruine de Bruin (2010) found that people often systematically overestimate the energy used by highly salient but low-energy activities, such as having the lights on. In reverse, people systematically underestimate the energy used by high-energy appliances that are used less frequently (e.g., the washing machine).

There are several reasons why people know little about their domestic energy use. First, energy is invisible (Chisik, 2011; Maréchal & Holzemer, 2015; Schwartz, Denef, et al., 2013). When switching on a device, it is not transparent how much electricity (or gas) is consumed. Second, as suggested by social practice theory (Hargreaves, 2011), people do not seek to use electricity, but they use electricity to satisfy a need and accomplish a goal such as cooking or washing laundry (Entwistle et al., 2015; Rego Teixeira, 2014). Third, conventional energy bills have typically summarised energy usage over an extended period of time (usually one or three months). This aggregated format is not usable for the consumer, as was illustrated by Kempton and Layne's (1994) analogy with a monthly grocery bill. Hence, households do not learn how much energy they consume for everyday household chores (Vanhulst & Lalanne, 2015).

Householders' limited knowledge about their energy consumption affects how they go about saving energy and impairs reasonable decision making in the process. There is consensus in the literature that, in line with social practice theory, feedback needs to address those very daily routines for householders to learn about their consumption patterns and make reasonable changes (Darby, 2001; Fischer, 2008; Sweeney, Kresling, Webb, Soutar, & Mazzarol, 2013). For interventions aiming to increase energy literacy, this means that feedback given to householders should prioritise practical information over technical information and it needs to be tailored to practices within the home and map to everyday activities to allow users to integrate it into their knowledge structures, i.e., their mental models of energy usage (Álvarez & Vega, 2009; Gabe-Thomas et al., 2016; Hofman, 1980; Palm & Ellegård, 2011).

2.4.4 Conclusions

First, energy literacy in householders is low, which adds a level of difficulty to residential energy feedback and chances are that users do not understand energy-related information. Second, there is inconsistency in the literature with regards to how energy literacy is defined and measured. Studies on energy feedback talk about increasing energy literacy, often without explicitly defining energy literacy for the scope of their work but merely referring to existing definitions. The publications by DeWaters (2011, 2013) are the predominant source for the definition of energy literacy and they provide the only validated questionnaire to measure it. While field studies often adopt the definition, they do not use the validated questionnaire. The questionnaire does not seem very useful for energy customers, as it was designed for school students. The other challenge with the questionnaire is that it conflates knowledge, attitude, and behaviour (Mogles et al., 2017).

2.5 Graphical Literacy

Energy data can be displayed as numeric information in units of energy (kilowatts) or costs (monetary units). Often, it is fed back to users in graphical format or in abstract visualisations. Visualising energy data is considered important as a way to assist householders in reducing energy consumption (Murugesan, Hoda, & Salcic, 2014).

This section reviews definitions, strengths and weaknesses of graphs and other data visualisations and ways how user's comprehension of graphs and visualisations can be measured.

2.5.1 Definition

Data visualisation (Data Vis) is the presentation of data in a pictorial or graphical format. Going beyond basic graphs and charts, research has evolved around complex visualisation systems. Information visualisation (Info Vis) often refers to computer-based systems that help domain experts explore or explain data through interactive software that exploits the capabilities of the human perceptual system (Munzner, 2014). The term visual analytics emphasises a problem-solving process where analytical reasoning is supported by an interactive visual interface (this may include sophisticated algorithms and computations to create the visualisation). *Casual Info Vis* refers to depictions of data in everyday life as opposed to many Info Vis systems that are typically used by domain experts (Pousman, Stasko, & Mateas, 2007). Casual Info Vis can include mobile and ubiquitous interfaces and deals with personally relevant data that might be visualised in ambient and artistic ways. Balchin and Coleman (1966) coined the term *graphicacy*. They define it as:

> *'the intellectual skill necessary for the communication of relationships which cannot be successfully communicated by words or mathematical notation alone; it is a skill to be possessed by both those wishing to communicate and those attempting to understand visual aids, especially maps, photographs, charts and graphs, are the media of communication'.*

Later works use the term *graphical literacy* which is defined as '*the ability to read and write (or draw) graphs'* (Fry, 1981). Cleveland and McGill (1984) define *graphical perception* as '*the visual decoding of information encoded on graphs'*. According to Pinker (1990), *graph comprehension* relies on perceptual and cognitive components to identify relevant pictorial elements and relate them to the real-world matter that the graph depicts. Similarly, Boy, Rensink, Bertini, and Fekete (2014) talk about *visualisation literacy*, which they define as the ability to use a given data visualisation to translate questions in the data domain into visual queries in the visual domain, and to interpret visual patterns in the visual domain as properties in the data domain.

2.5.2 Benefits

Graphical representations often communicate data better than textual representations because they support human cognition in processing quantitative information (Fry, 1981; Larkin & Simon, 1987; Pinker, 1990). If the single most important aim was to display numeric data in its most accurate form, lists and tables would be preferable (Tufte, 1983). However, often one wants to trade the accuracy of quantitative data for a more accessible and engaging visualisation. External representations enhance thinking by saving internal memory and providing structure (Kirsh, 2010; Munzner, 2014) and therefore they play a critical role for human cognition, problem solving and conceptual learning (Cheng, Lowe, & Scaife, 2001).

If characteristics of graphs are suitable for human perception and cognition, it is easy to correctly decode the information (Cleveland & McGill, 1984). Suitable representations can significantly enhance understanding in complex domains, determine what is learnt and how easily and quickly it is learnt. Pinker (1990) proposes that people have general schemas containing knowledge of what graphs are and how to read and interpret them. The idea of schemata reoccurs in Zhang and Norman's (1994) principle that representations should always match the physical properties of what they represent. It is also in line with Cheng's (2014) 'Representational Epistemic' approach which *'claims that the key to understanding the efficacy of a notational system (…) is to focus on how the specific representational schemes (…) of a notation encode the core concepts that permeate a knowledge domain'*. That is, the representational structure should preserve the conceptual structure of people's mental model of the problem (Cheng & Barone, 2017). If the representation matches the mental model, people understand and learn more easily (Cheng, 2011). In the context of this thesis, this means that energy data visualisations need to be designed in a way that will support understanding and learning.

2.5.3 Challenges

Visualising data for everyday life seems a good idea. However, people tend to feel overwhelmed by the vast number of information technologies these days. The flood of information can be perceived as intrusive and stressful (Carpendale, 2013). It is therefore important to choose a user-centred design approach for the presentation of data to make it as accessible and easy as possible for the user.

While suitable graphs enhance cognitive analysis, poorly chosen graphs decrease the ease with which users make sense of them (Baur, Lee, & Carpendale, 2012; Tufte, 1983). In a report from the Centre for Sustainable Energy (an independent national charity) to Ofgem (the government regulator for gas and electricity markets in Great Britain), it says that *'the manner of presentation of the feedback information to consumers is a core consideration which has been much overlooked in the literature'* (Roberts & Baker, 2003). This is problematic, considering that energy information is often visualised graphically and Galesic and Garcia-Retamero (2011) suggest that even the simplest graphs may be difficult to understand for many people.

Peebles, Ramduny-Ellis, Ellis, and Bonner (2013) investigate the influence of graph schemas, defining schemata as 'knowledge structures' that represent graphical and representational properties of diagrams, as well as the knowledge of how to interpret them. They presented unfamiliar diagrams to participants and found that people use familiar diagrams' schemata to interpret them. The more the representational features of the presented diagram resemble the features of the familiar diagrams that are being used as reference, the more accurate is the interpretation. Where people make false interpretations, the errors are characterised by biases that result from people applying familiar diagram schemata to unfamiliar diagrams that do not share the same features.

2.5.4 Measures

Graph comprehension mostly refers to people's ability to read and interpret graphs (it can also mean the skill of choosing, constructing, or inventing graphs). Friel, Curcio, and Bright (2001) establish that graph comprehension involves the three levels *translation, interpretation, and extrapolation/interpolation*. This corresponds to Bertin and Barbut's approach from 1973, which proposes the three interpretation levels: elementary, intermediate, and comprehensive. The elementary or translation level is to translate a graph into a semantic description, i.e. a person would be able to extract information, explain verbally the specific structure of the graph and to interpret it at a descriptive level. The intermediate or interpretation level goes a step further and requires the understanding which factors are most important. This may involve understanding relationships, for example between the lines in a graph and the according labels of the axis. The comprehensive or extrapolation and interpolation level include extended interpretations and inferences, meaning one not only interprets the graph but understands consequences, such as spotting trends or extrapolating further implications. These three levels build on the concept of literacy, which is equated with using written information (both reading and producing) to function in society.

Comprehension questions requiring different depths of analysis have been applied in practical research to assess people's graph comprehension. For example, Galesic and Garcia-Retamero (2011) tested a catalogue of questions with two large samples to develop a test scale to measure health-related graph literacy. They presented a line graph with time (years from 1970 to 2005) on the x-axis, and percentage of people with a certain disease on the y-axis. Amongst the comprehension questions they asked were the following: Approximately what percentage of people had the disease in the year 2000? When was the increase in the percentage of people with the disease higher – from 1975 to 1980 or from 2000 to 2005? According to your best guess, what will the percentage of people with the disease be in the year 2010? These three questions represent items from the three levels translation, interpretation, and extrapolation. The first question simply requires participants to read off a point on a line chart. The second question requires them to compare slopes of a line at two intervals. The third one requires them to project a future trend from the line chart.

2.5.5 Critique of Measures

Measuring graph comprehension with standardised measures is challenging since graphs come in so many variations and because every case conveys different information. Therefore, measures and semantic questions to test people's understanding must be tailored to the specific scenario. Particularly relevant for the case of energy data visualisations is whether householders understand the implications for everyday life and extrapolate from the data feedback to consider how changing behaviour would change the data. Even though information on energy consumption is quantitative in nature, the measure to assess householders' faceted understanding of energy data may be best captured with a set of qualitative questions.

2.5.6 Conclusions

With ubiquitous computing, more and more data are becoming available to end-users. It is therefore crucial to explore the potential that visualisations and design principles have in everyday domains, such as residential energy feedback. It is important to visualise information in a comprehensible way that relates to people's everyday social practices. The way data is presented – whether it is visualised as numbers, graphs, or other abstract representations – is crucial for people to obtain understanding of their home energy usage data.

2.6 Disaggregation

The previous sections have established that householders have poor knowledge about how they use energy and that feedback needs to map to everyday practices in the home to be useful to householders. Most sensing infrastructures that are used in smart energy feedback today, use only one central sensor in the home to record the household's total energy consumption. This approach does not typically provide data on how much energy is consumed by individual appliances. This section reviews the opportunities to obtain appliance-centric data.

2.6.1 Approaches to Disaggregating Energy Data

A household's total electricity usage data can be broken down into the consumption of individual appliances. There are two possibilities for obtaining consumption information at the level of the individual appliance. The first one is to have one sensor per appliance, which collects data directly at the device-level when it is used (e.g. Froehlich, 2011; Kelly & Knottenbelt, 2015). This solution is straightforward and accurate, but neither convenient nor time- or cost-efficient. It works for small-scale studies but it is not a solution in nationwide rollouts. Given the practical challenges of having a sensor associated with every appliance in the home, research has explored alternative approaches to determine how much energy individual appliances are using in the home.

Non-Intrusive Load Monitoring (NILM) is an approach to determining the amount of electricity used by individual appliances in the home that uses mathematical algorithms to disaggregate unique signals detectable in data read from a single sensor monitoring the household's total consumption at one point in the household. Approaches to NILM date back to the 1980s (Hart, Kern, & Schweppe, 1989), and techniques have been further advanced in recent years (Armel et al., 2013; Goncalves, Ocneanu, Berges, & Fan, 2011; Patel, Robertson, Kientz, Reynolds, & Abowd, 2007). Current NILM algorithms are not very exact and no solution is suitable for all types of appliances (Zeifman & Roth, 2011). NILM research is often hard to evaluate and to compare because every study features different data sets, algorithms, metrics, resolutions and accuracy (Batra et al., 2014; Reinhardt et al., 2012). Especially when several appliances are running simultaneously it becomes harder both for the human eye as well as for machine learning to identify their signatures and disaggregate them from the total usage pattern.

2.6.2 Benefits

The obvious benefit of disaggregation is that it simplifies the complex nature of domestic energy usage data. Environments are considered *complex* when they involve different types of entities, each with several properties, which again can take several different values (Cheng & Barone, 2017; Osman, 2011). In the context of energy use, this would be, for example, numerous devices that use energy, and they can be used in different ways, impacting on how much energy is used. For example, the washing machine has different programmes, and each of these programmes vary in the amount of energy that they consume. Therefore, it is not possible for user to simply ascertain how much energy washing a load of laundry uses, as it depends on which programme is selected.

Householders as ‘problem-solvers’ (Cheng & Barone, 2017) add complexity because of their varying degrees of abilities and expertise. Different members of the same household use energy in different ways, and one member does not necessarily know what the others are doing. Their abilities to analyse feedback might vary, too, and understanding energy usage patterns without knowing what someone else has done is difficult even for someone with good analytical skills. Finally, autonomous changes in appliances’ power usage add another layer of complexity to home energy use. For example, refrigerators do not have a stable power consumption, but go through cycles when they cool down and then stop cooling until the temperature reaches a certain cut-off again. Humans are biased towards attributing changes in the feedback to their own actions (which are more salient) and neglect autonomous changes (which are invisible and unknown to the average user) (Osman, 2011).

There is consensus in the literature that disaggregation is a desirable feature for energy feedback. Neustaedter, Bartram, Mah (2013) write that residential energy feedback *‘cannot be simply presented in aggregate’*. The assumption that disaggregated data should be more useful has high face validity, because total consumption provides no link to the impact of everyday household practices and routines and thus lacks relevance with regards to social and situational factors. Without this, feedback will be less meaningful for householders and cannot trigger sufficient reflection or learning.

Álvarez and Vega's (2009) publication on environmental education discusses the relationship between environmental concern and responsible environmental behaviour. What they describe reflects the barrier between Reflection and Action in Li, Dey, and Forlizzi's (2010) model of personal informatics (2.2.1). Despite awareness and concern about the environment, people might not behave in an environmentally-friendly way. Álvarez and Vega see the challenge for environmental education to overcome the gap between theoretical discourse and everyday life and they outline didactic guiding principles for interventions. One of these principles is that people need to be enabled to move from knowledge to action: they must learn how to diagnose and analyse everyday situations in their life and plan specific activities that they want to change.

2.6.3 Challenges

Studies on disaggregated eco-feedback have brought forward inconsistent results. Several studies have found evidence in favour of appliance-specific data (Froehlich, 2011; Hargreaves et al., 2010; Schwartz, Denef, et al., 2013; Wood & Newborough, 2007). These studies emphasise the added value, arguing that disaggregated data is richer in information and has therefore more potential to empower homeowners. Fischer, 2008 found in her review of five review studies over 21 original studies that appliance-wise disaggregation helps consumers detect how much energy appliances consume, which gives them a sense of control to change their use of these appliances.

In a more recent review study, Kelly and Knottenbelt (2016) reviewed eleven field studies on disaggregated energy feedback and concluded that they do not provide evidence in favour of disaggregated feedback. At the same time, they point out that all reviewed studies suffer from methodological biases. For example, Sokoloski (2015) found that aggregated feedback lead to more saving than disaggregated feedback but the findings were confounded by how often participants had engaged with the feedback. The aggregated feedback was displayed on an IHD and disaggregated feedback was provided through an online platform. Participants accessed the web-feedback far less than the other group looked at their IHD. Based on the current state of the art it is difficult to evaluate whether disaggregated data feedback has an advantage over non-disaggregated feedback in increasing householders' knowledge about their energy use and in decreasing their consumption.

2.6.4 Conclusion

Appliance-level information would help to break down the complexity of residential energy use. Single sensors are not a practical solution to collect this data and algorithms to extract it from the total consumption are not working well enough yet. Also pending is conclusive empirical evidence to demonstrate how much value disaggregation would bring to energy feedback. Given this lack of evidence, this brings up the question whether there is a clear case for research to continue to invest resources into developing solutions.

2.7 Conclusions from the Literature Review

Smart energy feedback has the potential to better inform householders about their consumption than conventional feedback in the form of monthly bills. This review of the current literature around smart energy feedback has revealed several challenges that are keeping these devices from being as effective as they could be in helping people to reduce their consumption.

A critical question, and the main RQ in this thesis, is whether householders can easily understand smart energy feedback. Their understanding or learning is often referred to as energy literacy. One gap identified in the literature review is that the term energy literacy is widely used to explain how energy feedback is meant to increase knowledge, and in consequence, to decrease consumption. A definition of energy literacy has been offered by DeWaters and Powers (2011) and publications on energy feedback often refer to this definition. However, this definition is relatively broad, and conflates knowledge, attitude, and behaviour, and so does the energy literacy questionnaire (DeWaters et al., 2013). Moreover, the questionnaire is aimed at high-school students rather adult householders. RQ1 addresses this discrepancy: What is energy literacy in the context of home energy use?

For householders to increase their understanding about energy, they need to understand the information that is given to them. A lot of research has examined the effect of how feedback is given on energy savings. There is less research investigating in depth how users reflect on feedback and whether it increases their understanding of how they are consuming energy through their daily activities. Reflection and integration are considered crucial steps in theoretical models that explain behaviour change in the context of persuasive technologies. Despite the frequent use of visuals and graphics in home energy feedback, there is little research on how graphical literacy impacts householders' ability to analyse them. RQ2 builds on this limitation of existing research: How do householders interact with smart electricity feedback? Can they make sense of it? Do they understand it?

RQ3 - How does the design of the data visualisation affect how people make sense of domestic energy data? - builds on the literature review on energy literacy, graphical literacy, and disaggregation, but also on findings from the studies that were done to address the first two research questions. The main reason this thesis focuses on electricity feedback (not energy including gas), is that disaggregation has been a central focus in energy feedback research for the past decade. The multitude of electric household appliances makes electricity the more interesting use case for disaggregation, despite gas constituting the bigger consumer of energy in UK homes.

Chapter 3 Defining Energy Literacy

3.1 Introduction to Study 1

Study 1 addresses RQ1 of this thesis: What is energy literacy? The literature review revealed inconsistent use of the term energy literacy. Firstly, energy literacy has been defined as including sound knowledge about energy, attitude towards it, and energy-related behaviour (DeWaters et al., 2013). This definition conflates cognitive, attitudinal and behavioural components (Ajzen & Fishbein, 1977; Mogles et al., 2017). Second, numerous studies rely on DeWaters' work, without addressing that it has been developed in a high-school context, where pupils received physics education, and then took a test to assess their energy literacy. Smart energy feedback in the home is very different from physics classes in school, hence the question arises: how does DeWaters' work apply to the context of home energy use?

From a psychometric point of view, it is necessary to first define what energy literacy is for the scope of investigating smart energy feedback in the home. Based on the information-deficit model (Dickson, 2005), the implicit assumption of smart energy feedback is that householders become more informed (more energy literate) and as a result reduce their consumption. Existing field studies often rely on DeWaters' definition, but they use their own questions to assess energy literacy in householders, instead of using the inventory developed by DeWaters et al. (2013) or the adapted questionnaire by Brewer (2013). DeWaters' and Brewer's inventories, as well as questionnaires used in home energy studies, include technical questions, such as items that require respondents to calculate energy consumption. The questions asked in field studies of home energy use go beyond this and include questions that are practical in nature, for instance, they ask how much householders pay for their energy bills (e.g. Brounen et al., 2013).

There is a discrepancy between the available definitions and validated measures, and the need of home energy studies for an actionable definition and user-friendly measures of energy literacy. The exact content of questions to measure energy literacy may vary between use cases and studies. The biggest conflict however, seems to lie in the definition of energy literacy as either encompassing cognitive, attitudinal, and behavioural elements, versus literacy equalling cognitive components only relating to what householders know about energy use in their home.

The purpose of Study 1 is to examine definitions and measures of energy literacy and to define energy literacy and ways to assess it for the scope of this thesis. The method chosen to do this were focus groups (Krueger & Casey, 2014; Morgan, 1997; Vaughn, Schumm, & Sinagub, 1996). Focus groups are guided group interviews with the purpose of listening to participants to learn from them while they discuss a given topic. The data is qualitative in nature and provides a deep understanding of participants' experiences and opinions with regards to energy literacy. Three focus groups were conducted with the following groups: energy experts from academia (academics working at UCL's Energy Institute with backgrounds in physics and engineering), energy experts from industry (employees of EDF Energy UK), and householders (lay group). This allowed us to conduct a participant-based group analysis (analysing individual contributions in every group) as well as a whole group analysis (treating the data of a group as one and comparing it to the other groups) (Ritchie, Lewis, McNaughton Nichols, & Ormston, 2014). The data analysis followed the framework analysis approach, which contains interconnected stages in which the author familiarised herself with the data, highlighted important passages and sorted quotes, made comparisons within and between cases, and interpreted the findings (Rabiee, 2004; Richie & Spence, 1994). All focus groups were audio recorded and transcribed in the transcription software f5. The transcripts were coded and iteratively analysed in Word MS Office.

3.2 Method

We conducted three focus groups with a total of 20 participants. The sessions included activities and group discussions which are further described in Material (3.2.2) and Procedure (3.2.3). The data was analysed both on the individual and the group level. Similarities and differences between the three groups were analysed on the group level without delineating individual contributions. Within the groups, we follow a participant-based group analysis, i.e. the contributions of individual participants were analysed within the context of the group discussion. The analysis followed the themes defined by the focus group procedure (top-down analysis), but also allowed themes to develop from the narrative of participants' discussions (bottom-up analysis).

3.2.1 Sample

The first focus group was conducted with seven academic energy experts from UCL's Energy Institute. The second focus group was conducted with seven energy users, we used a convenience sample consisting of students and administration staff from UCL (who did not study or work in energy). The third focus group was conducted with six energy experts from industry, namely employees of EDF Energy UK. The three groups varied in age – group one was the most diverse ranging from PhD students in their mid-twenties to senior professors. The second group was composed of students and administrative staff members of UCL in their twenties to mid-thirties. The EDF employees in group three were in their mid-thirties to mid-forties.

3.2.2 Material

During the focus groups, we used PowerPoint slides projected onto a screen and printed handouts.

3.2.2.1 Slides

The slides were used to guide the session. They contained an overview of the agenda, the two lead questions for the focus group and references from the scientific literature on energy literacy and measures of energy literacy. The two lead questions were:

1) *What is energy literacy?*

2) *How can we measure energy literacy?*

The extracts from the literature were definitions of energy literacy, statements that related to practical knowledge and behaviour with regards to energy use in the home, and measures of energy literacy. The energy literacy definitions were the following:

'Energy literacy is a broad term encompassing content knowledge as well as a citizenship understanding of energy that includes affective and behavioral aspects' (DeWaters & Powers, 2013).

'An energy literate person needs to have a basic understanding of energy concepts. A sound knowledge base is important' (DeWaters, Powers & Graham, 2007).

'Energy literacy is the understanding of energy concepts necessary to make informed decisions on energy use at both individual and societal levels. Increasing energy conservation is difficult when people do not understand energy fundamentals, or how energy is used in their homes and work-places. (…) Some examples of energy literacy are: Understanding the difference between power and energy. Knowing that a microwave uses much more power than a refrigerator, but that the refrigerator will use much more energy over time. Knowing how electricity is generated in one's community' (Brewer, 2013).

The statements about practical knowledge and home energy use behaviour were:

'Most people have only a vague idea of how much energy they are using for different purposes and what sort of difference they could make by changing day-to-day behaviour or investing in efficiency measures' (Darby, 2006).

'Feedback should prioritize practical knowledge over technical knowledge' (Hofman, 1980).

'Interventions need to provide the practical mapping of information to everyday activities' (Álvarez & Vega, 2009; Ellegård & Palm, 2011).

'Energy-related feedback must be tailored to practices within the home' (Gabe-Thomas, Walker, Verplanken & Shaddick, 2016).

The overview over the validated energy literacy inventories were the following two bullet points:

'affective, cognitive and behavioral scales' (DeWaters , Qaqish , Graham & Powers, 2013).

'attitude, knowledge and behavior scale' (Brewer, 2013).

An additional slide listed other measures that have been used in the literature to assess energy literacy:

> *'Ranking task according to household appliances' energy consumption' (Yun et al., 2010).*
>
> *'Power-rating quiz of different household appliances' (Anderson & White, 2009).*
>
> *'Pairwise comparisons' (ENLITEN game,* www.cs.bath.ac.uk/enliten).

3.2.2.2 Handout

The first page of the handout was a blank sheet for participants to write on during Activity I (3.2.3). The following pages of the handout were the items from the energy literacy questionnaires by Mogles et al. (2017), Brounen et al. (2013) and Brewer (2013).

The questionnaire by Mogles et al (2017) included the following six questions:

1. *How much do you feel you know about energy? A lot (expert)/Quite a bit (informed)/Not much (novice)/Nothing*
2. *Which of the following sources of information has contributed most to your understanding of energy issues? Further or higher education /School/Books newspapers or magazines/ Friends or family members (including parents)/ Internet/Television/iBert system [used in the study]/Other (please specify)*
3. *The term renewable energy resources means? Resources that are free and convenient to use/Resources that can be converted directly into heat and electricity/Resources that can be converted directly into heat and electricity/Resources that do not produce air pollution/Resources that are very efficient to use for producing energy/ Resources that can be replenished by nature in a short period of time*
4. *Most of the renewable energy in the UK comes from which of the following sources? Solar/Water (hydro/tidal/wave) power/Wind/Landfill gas/Geothermal/Don't know*
5. *Which of the following actions, if everyone did this all the time, would save the most energy in the UK? Turn off lights when they are not in use/Turn down the heat in rooms/Reduce water consumption/Walk or cycle short distances instead of going by car/ Turn appliances off at the plug*
6. *Which kind of lighting uses the least amount of energy? Standard light bulbs/Low energy light bulbs/Fluorescent lights/LED lights/Don't know*

The questionnaire by Brounen et al. (2013) included the following four questions:

1. *How much do you pay for your monthly gas (electricity) bill? A) ___ euro B) I have no idea.*
2. *Suppose you own your home, your heating system breaks down and is beyond repairs. As a replacement, you can choose between two heating systems. Model A is for sale for €3,750 and is expected to result in a monthly gas bill of €100. Model B is more expensive, with a retail price of €5,000, but will result in a monthly gas bill of €80. You can assume that both models have an economic lifespan of 15 years. Which heating system do you prefer? Heating system A/Heating system B/I have no preference, both models are equally adequate/I have no idea*
3. *At which temperature do you set your thermostat during the evening? ___ degrees Celsius.*
4. *At which temperature do you set your thermostat at night? ___ degrees Celsius.*

The questionnaire by Brewer (2013) was chosen instead of the original questionnaire by DeWaters et al. (2013) as DeWaters et al.'s publication provides only the questions with the correct answers in the appendix but without the multiple-choice options. Brewer's adapted questionnaire is accessible with all response options. Items marked with an (R) were reverse scored.

The attitude scale includes the following 18 items to be responded to on a five-point Likert-style scale from *Strongly agree* to *Strongly disagree*, plus *Choose not to answer*:

1. *Energy education should be an important part of every school's curriculum.*
2. *I would do more to save energy if I knew how.*
3. *Saving energy is important.*
4. *The way I personally use energy does not really make a difference to the energy problems that face our nation. (R)*
5. *I don't need to worry about turning the lights or computers off in the residence halls, because the school pays for the electricity. (R)*
6. *Americans should conserve more energy.*
7. *We don't have to worry about conserving energy, because new technologies will be developed to solve the energy problems for future generations. (R)*
8. *All electrical appliances should have a label that shows the resources used in making them, their energy requirements, and operating costs.*
9. *The government should have stronger restrictions about the gas mileage of new cars.*
10. *We should make more of our electricity from renewable resources.*
11. *America should develop more ways of generating renewable energy, even if it means that energy will cost more.*

12. *Efforts to develop renewable energy technologies are more important than efforts to find and develop new sources of fossil fuels.*
13. *Laws protecting the natural environment should be made less strict in order to allow more energy to be produced. (R)*
14. *More wind farms should be built to generate electricity, even if the wind farms are located in scenic valleys, farmlands, and wildlife areas.*
15. *More oil fields should be developed as they are discovered, even if they are located in areas protected by environmental laws. (R)*
16. *I believe that I can contribute to solving the energy problems by making appropriate energy-related choices and actions.*
17. *I believe that I can contribute to solving energy problems by working with others.*
18. *Many of my everyday decisions are affected by my thoughts on energy use.*

Brewer's behaviour scale includes the following 17 items to be responded to on a five-point Likert-style scale from *Always or almost always* to *Never or hardly ever*, plus *not applicable*:

1. *I turn off all appliances (TV, computer, game console, etc) every night before going to sleep.*
2. *I leave my computer and/or monitor on, even when they are not being used. (R)*
3. *I turn off vampire loads (like cell phone chargers) using a power strip.*
4. *I leave the lights on when I leave a room. (R)*
5. *I use task lighting (like desk lamps) rather than overhead lighting.*
6. *I use sunlight rather than electric lighting whenever possible.*
7. *I take the stairs rather than the elevator whenever feasible.*
8. *I drive alone (no passengers). (R)*
9. *I walk, bike, or roll to go short distances, instead of driving.*
10. *I use public transportation.*
11. *I recycle my cans and bottles.*
12. *I bring reusable bags when shopping.*
13. *I eat meat. (R)*
14. *I turn off water when brushing my teeth, shaving, etc.*
15. *I turn off water in the shower when soaping and scrubbing.*
16. *I wash only full loads of laundry.*
17. *I wash my laundry in warm or hot water. (R)*

Brewer's knowledge scale includes the following 13 multiple-choice items:

1. *Electrical power is commonly measured in units of: volts (V)/watt-hours (Wh)/joule (J)/watts (W)/British Thermal Units (BTU)/Choose not to answer*

2. *What is the primary cause of current climate changes? Carbon dioxide released from burning fossil fuels/There is no cause, climate change isn't real/Natural solar cycles/Radioactive waste from nuclear power plants/Melting glaciers in Greenland/Choose not to answer*
3. *Electrical energy is commonly measured in units of? Erg/ampere (A)/British Thermal Units (BTU)/watt-hours (Wh)/watts (W)/Choose not to answer*
4. *What is the breakdown of the clean energy mandated by 2030 by the Hawaii Clean Energy Initiative? 20% from renewable sources, 80% from energy conservation/30% from energy conservation, 40% from renewable sources/50% from renewable sources, 10% from conservation/30% from solar, 30% from wind, 10% from waves/30% from renewable sources, 20% from conservation, 10% from natural gas/Choose not to answer*
5. *Order these types of light sources from lowest to highest power usage, assuming they provide the same amount of light: incandescent bulb/compact fluorescent lightbulb (CFL)/light-emitting diode (LED)*
6. *Approximately how much carbon dioxide (CO_2) is in the atmosphere now, and what level is considered the safe upper limit to avoid the worst effects of climate change? 450 ppm CO_2 in atmosphere now, 500 ppm CO_2 safe upper limit/331 ppm CO_2 in atmosphere now, 350 ppm CO_2 safe upper limit/393 ppm CO_2 in atmosphere now, 350 ppm CO_2 safe upper limit/600 ppm CO_2 in atmosphere now, 450 ppm CO_2 safe upper limit/100 ppm CO_2 in atmosphere now, 50 ppm CO_2 safe upper limit/Choose not to answer*
7. *Order these appliances from lowest to highest power usage: desk lamp with compact fluorescent lightbulb (CFL)/mobile phone charger (while charging)/plasma TV/microwave/laptop*
8. *On average, how much electrical energy does a home in Hawaii use per day? 400 W/20 kWh/87 kWh/328 kWh/4kWh/Choose not to answer*
9. *What is the approximate maximum power generated from a single standard rooftop solar panel? 25 W/800 W/50 W/10 kW/200 W/Choose not to answer*
10. *What are the expected long-term effects of current climate changes? A significant rise in the sea level/Global temperatures increasing by a few degrees on average/Increasing sea water acidity/Changes in seasonal rainfall patterns (droughts, floods)/All of the above/Choose not to answer*
11. *What is currently the source of approximately 80% of Hawaii's electricity? oil/wind/natural gas/coal/solar/Choose not to answer*
12. *A compact fluorescent lightbulb (CFL) uses 13 W. If it is run for 2 hours, how much energy does it use? 13 Wh/7.5 Wh/26 Wh/130 Wh/52 Wh/Choose not to answer*
13. *If your game console uses 200 W when turned on, how much energy would it waste if you left it on all weekend while you were away? 15000 Wh/100 Wh/960 kWh/9.6 kWh*

There were minor changes in the materials between groups. In the first focus group with the academics, we used the term research questions on the slides. This was changed to simply questions for the second and third focus groups with users and industry. Another change concerned the material with regards to the energy literacy questionnaires: only Brewer's questionnaire was handed out on paper in the first focus group with the academics; the questions by Brounen et al. were presented on the slides and the questions by Mogles et al. had not yet been published. For the latter two groups (users and industry), we added the questions by Brounen et al. and Mogles et al. to the handout.

3.2.3 Procedure

The focus group sessions comprised an introduction by the researcher, two activities and group discussions. The focus groups were structured as follows:

Introduction:	*Welcome and introduction to the focus group by the researcher*
Activity I:	*What is energy literacy? Instruction: Think about the term 'energy literacy' for two minutes and what it means to you. Write down three features that define energy literacy in your opinion.*
Discussion I.I:	*Discussion of Activity I*
Discussion I.II:	*Presentation of extracts from the literature defining energy literacy and addressing practical knowledge and behaviour with regards to home energy use.*
Discussion II.I:	*How can we measure energy literacy? Presentation of energy literacy inventories from the literature.*
Activity II:	*Instruction: Please go through the energy literacy questionnaires in the handout and tick all questions that you think are suitable, and cross out the questions that you think are unsuitable to test someone's energy literacy.*
Discussion II.II:	*Discussion of Activity II. Presentation of additional (non-validated) measures of energy literacy.*
Summary:	*Researcher's summary of the focus group discussions and invitation of participants to confirm if the summary is accurate and to add to it.*
Closing:	*Room for final comments and questions from participants.*

The introduction outlined the purpose and rules of focus groups, i.e. it was explained that there were no right or wrong opinions, and that everyone's perspective mattered. The focus group started with Activity I. The session started immediately with the first activity to not influence participants by showing definitions of energy literacy from the literature first, but to explore what participants intuitively thought energy literacy means. Participants were provided with a blank sheet and pens and given a few minutes to silently write down three key aspects of what they thought defines energy literacy. Once everyone signalled they were done, the researcher invited the group to present and discuss what they had written (Discussion I.I). The researcher then presented slides with definitions of energy literacy and statements about knowledge and behaviour with regards to energy use in the home (see Material). In Discussion I.II, participants discussed the provided literature, compared it to their own definitions and debated whether they agreed with the literature. This first part of the focus group addressed the lead question 'What is energy literacy?'.

Discussion II.I introduced the second lead question: How can we measure energy literacy? To that end, the researcher presented a slide introducing the three energy literacy scales for cognition, attitude, and behaviour. Participants briefly discussed what they thought about the three scales. This was followed Activity II, which involved participants reading the handouts of energy literacy questionnaires (see Material). Once every participant finished reading, they discussed the suitability of the questionnaires' items (Discussion II.II). When time allowed, the researcher provided an additional slide that listed alternative energy literacy measures, namely power quizzes and ranking tasks. At the end of the focus groups, the researcher summarised everything that the group had talked about and invited them to confirm her summary or correct it and to add to it. Participants were given the opportunity to make final comments and ask questions before the group separated.

The focus groups lasted a little over an hour. The content was the same between the three focus groups, other than the small differences pointed out in the Material section. Due to the group dynamics, the timing varied slightly between the groups and not all slides were given the same attention between the groups. Similarities and differences are described in the results and addressed in the discussion.

3.3 Results

The results are presented in order of the activities and discussions in the focus groups. We first describe our analysis of participants' written statements from Activity I, dissecting their definitions of energy literacy and putting them into context with what was said during Discussion I. This is followed by the analysis of the second half of the focus group which addressed the question of how to measure energy literacy (Activity II and Discussion II). The letters A, U and I denominate academic, user and industry participants, the digits enumerate the participants per group (A1-A6, U1-U7, I1-I6).

3.3.1 What is Energy Literacy?

3.3.1.1 Activity I

In Activity I, participants were asked to write down their own spontaneous definition of energy literacy. A full list of all definitions written down by participants during Activity I can be found in Appendix A. The biggest overlap of participants notes would be the following definition:

> *'Understanding how energy is used, where it comes from and what the implications are'.*

How energy is used includes a practical understanding relating to everyday use for activities, habits and appliances or processes. Where energy comes from relates to the generation of energy as well as its transmission via the grid. The implications our participants referred to concern the CO_2 emissions and their global environmental impact associated with different types of energy generation, and hence the societal need to use energy sustainably and reduce waste.

Two participants (A2, U4) named understanding common energy units (kilowatts and kilowatt-hours) as an aspect of energy literacy which relates to understanding the bill (U4) and one academic thinks energy literacy could include the ability to deal with energy services (A6). The single most featured word in participants' notes is *understanding* with a frequency of 21 counts. All but one participant (I1) used either the word *understanding* (13 out of 20 participants) or the closely related terms *knowing* or *knowledge* (4/20), *interpret/make sense* (1/20) or *considering* (1/20).

Further, the verbs *decode* and *reflect* as well as *apply*, *express* and *describe* were used. Four participants, all from the users group, mention behavioural aspects in their notes:

> *'Perhaps also use with the goal of 'least consumption' and sustainable use' (U3).*
>
> *'make changes to behaviour to reduce energy consumption' (U4).*
>
> *'practical knowledge -> act, change behaviour' (U5).*
>
> *' Also about controlling it [energy consumption]?' (U6).*

3.3.1.2 Discussion I.I

In Discussion I.I, participants within one focus group read out what they had written in Activity I and discussed their definitions in the group. The results from this discussion are presented separately for the three focus groups.

3.3.1.2.1 Academics

A1 elaborated on her definition saying that she meant '*understanding how energy is used'* almost in a physical sense but also a practical understanding. The practical understanding would be knowing what the major energy uses in a home or in daily life are. She said for the practical understanding, one does not need to understand physics but '*know that a flight to Australia or heating your house in winter is much more significant than charging your mobile phone'*. There was consensus from the group that people should know that energy '*lights the bulbs and runs the machines'* (A7), even if they can't explain the exact physics behind it.

As A2 put it, there are different levels of energy literacy and one would expect a different level of literacy from say a homeowner compared to people closer to decision making positions, like utilities. It was pointed out that the questions *How is energy consumed? How is energy produced?* would be answered differently by different people but they agreed on the central role of energy for everyone ranging from comfort in life to national development and the economy. A2 and A3 worried that there was a risk with the assumption that if only people understood more, then they would do things differently, and taking a citizen view, we need a workable definition.

For the academics, the aspect of understanding of how energy is used would involve an understanding of wastefulness in the home. The aspect of where energy comes from would involve knowing if it has come from a wind turbine or solar panels for example. A2 pointed out that energy literacy is necessary but not sufficient to achieve behaviour change.

3.3.1.2.2 Users

In the users group, U1 elaborated on his definition saying a '*process or activity*' would be '*something like watching TV or boiling the kettle*'. U3 elaborated on her definition which included '*understanding and applying*', explaining that literacy means both having knowledge and being able to apply it. U4 said it included being '*able to read and understand your own energy bills*' and U7 emphasises the ability to '*interpret and make sense of energy data as for example in knowing what 400 kilowatts are and what that means, if certain appliances use that, is that a lot?*' Comparably, U6 talked about '*transferring skills*', meaning that if one knows what one machine consumes, can one apply to that to another context? U5 thought energy literacy '*entails theoretical and practical knowledge*'. He thought energy literacy includes the ability to reflect and decide, for example to adapt one's behaviour. Whereas U6 said that she put a question mark behind controlling energy consumption because she wasn't sure if that was part of energy literacy *or 'if that would be something different'*.

3.3.1.2.3 Industry

In the industry group, I4 read his definition:

> *'Understanding of the energy I use: where it comes from, what I use it for, how much I use'.*

Everyone in the group agreed, only I6 noted that he used the word *considering* instead of *understanding* because he thinks energy and knowing how it works is quite complicated and hence difficult to understand. Much like the academic groups, the industry group agreed that is it not necessary to understand the mechanical details of how, for example, a lightbulb works, rather, it's about context and knowing what an appliance can do for the user, how one uses energy and what it costs. From their perspective as utility employees, they would like customers to understand how they use their energy so that they would be responsive to services offered by the utility (e.g. load shifting). I4 added to the discussion that perspectives and values might matter, for example, if one was '*interested in the environment then the implication of my use of energy on the environment is important, but if I'm not interested, then it's not important'*.

3.3.1.3 Discussion I.II

Discussion I transitioned into its second part, where the researcher presented the definitions of energy literacy established in the literature (see Material). As before, the discussion of these definitions is presented separately for academics, users and industry in the following.

3.3.1.3.1 Academics

The academic group questioned the term *sound knowledge base* by DeWaters et al., not knowing what it is supposed to mean. They discussed an individualistic versus a social approach (e.g. '*simply heating because I am cold versus sitting around the fire place to spend quality time with the family'*). They challenged Brewer's assumption that literacy is necessary to make decisions, arguing that changes in society may introduce a new pattern or fashion. A1 suggested that when one studies science, one will know the difference between energy and power, which is useful, but how and why one is using energy is a matter of lifestyle and needs to be understood in the context of how society works. That is, one needs to think about the carbon emissions one has created by different actions such as using transportation or heating one's home.

The researcher then presented the statements about practical knowledge and behaviour with regards to home energy use (see Material). The academic group thought these '*made perfect sense'*, however, they also thought it was hard to provide information that is practical and maps directly to everyday actions. For example, if one received the instantaneous feedback of how much power the kettle is using when boiling water, that still would not tell the user what that means in terms of overall consumption. They agreed that practical knowledge that related to practices was more important than accurate technical knowledge, because after all practical knowledge can be good enough to get through life even if the underlying technical understanding might be flawed or wrong. Relating to how much one knows, A4 distinguished between *literacy* and *information*. He thought literacy is knowing that energy can come from coal or gas, or water, hydropower, or solar power, whereas knowing whether the energy one is receiving comes from coal (or not) is information which might be available (or not):

> *'in a very sort of simplistic silly way, it's like, it comes from coal rather than popcorn (…) and also you need to know the consequences of that. So, it comes from coal now, is that good or bad? Is it better than if it was coming from nuclear? Or is it worse?'.*

The group was hoping that feedback could politicise householders and make them more aware, or even alert them, to alternative (i.e. renewable) sources of energy.

3.3.1.3.2 Users

In the user group, the reaction to the energy literacy definitions by DeWaters et al. and Powers were slightly different. U1 was surprised, pointing out that he did not think one had to be knowledgeable about the physical properties and asking if a '*person who doesn't understand how electrons flow through a wire couldn't be energy literate?'* U5 was interested in the aspect of '*societal impact'* and explained that school had taught him only physics but that we need to learn sustainable practices. U2 thought there could be different levels of literacy and the expert level would include scientific knowledge. For the others, the debate of how much literacy is needed brought up the question of the purpose of energy literacy. There was consensus that the aim is to know where one is using energy and to be efficient in one's use:

> *'Maybe you take very short showers in the morning because you think it'll reduce energy when in fact it doesn't make a difference. Or maybe you keep*

the TV on very long because you think it won't make a difference but in fact it does?' (U7).

The group felt details about generation and transmission are less relevant to that end, however, U4 considered knowing the sources of energy - nuclear, coal, gas, wind, hydro-electric - as important for the political decision which providers to buy from. As for the three aspects behaviour, knowledge and attitude, U6 explained:

'Well, I could go "I have a hot shower every day. I know it consumes a lot of energy. I don't care"'.

U1 added:

'Or leaving the TV on standby rather than switching off at the plug. I know it consumes x amounts of kilowatts, x amounts of Pounds per year, but still my behaviour doesn't change because I don't wanna crawl into my media cabinet every single time I wanna watch TV'.

While U3 struggled to see the separation between the three, U2 analysed that:

'literacy is just knowledge. And it can change your behaviour or not. It's separate from knowledge. Just thinking about other things in life where I am literate but I can decide to ignore my knowledge'.

U1 added to that:

'Kind of like every smoker knows it's bad for your health but they still continue'.

U6, taking her peers thoughts into account, pondered if attitude might be the link between knowledge and behaviour:

'Attitude. Is that about making informed decisions?'.

The slides about practical knowledge and home energy behaviour were skipped in this focus group due to shortage of time.

3.3.1.3.3 Industry

The industry group, when presented with the energy literacy definitions from the literature, felt all definitions were right depending on how much depth or expertise one was expecting, with understanding fundamentals at the lower end of the gradient and domain expertise at the upper. I1 explained:

> *'I have a son who's nine and yesterday he came back and he said 'I've learned what's energy'. And it was a quick, a very simple way, to explain me how it was produced, where it goes, it goes through a cable, and I think for me that is energy literacy. Everybody knows at least, maybe not how it's produced in detail, but we're all conscious of energy, from very little we know it's not good to leave the lights on or use energy for nothing. There are different levels'.*

Talking about how literate their customers ought to be, they argued that customers should not need domain expertise at all, because EDF was doing the work for their customers. The only thing they would need to understand is how much energy they are using and what that's costing them. I6 brought up the following comparison:

> *'I don't know necessarily the technology behind me making a call or sending a message, or accessing the internet on my phone, I don't really need to know that, but I do need to know that if I send a picture message it costs me a Pound whereas if I go with my data, it doesn't cost me more money. So that's a level of literacy I have with my phone'.*

I2 took that comparison back into the energy context, saying:

> *'That's a really good example. When you switch (…) to flat rate, you've got your unlimited, so all of that care goes away, it doesn't matter. I don't know if you can equate that to energy but we talk about our green mix and moving towards renewables and stuff so that the importance of conservation and environmental impact and the cost in Pounds and Pence is going down and down and down and I think that there is a relationship between how much a customer needs to care about that and that singularity or whatever… some utopian future where we have endless green energy and zero cost to produce or something like that you know and we're all earning bitcoins, you know what I'm getting at. Then what's the need that the customer has to understand that, you know, it's zero, because it's not… it doesn't affect me at all'.*

Thinking more about the relationship of education, income, cost and literacy, I6 considered:

> *'You could have somebody who's well educated and on a good salary and they have no idea how much they use, so in terms of their personal consumption literacy zero, but they might understand the system that creates the energy. Somebody who is on low income, they have no understanding how it's produced but is super literate about the amount they use and how that affects them'.*

As for the question of whether attitude and behaviour are part of energy literacy, I5 disagreed with the nomenclature, saying literacy and behaviour are two different things but there was some back and forth in the discussion between the other participants. I2 elaborated that part of being literate in linguistics terms was being able to express oneself, and behaviour could be the equivalent in energy literacy. He conceded, though, when I5 reiterated that being literate did not necessarily change what one was doing. Like the user group, the industry group discussed that behaviour does not directly result from literacy, but it is mediated by attitude, or in I1's words, '*behaviour is the output in a way and knowledge and attitude are the inputs*'. At the end, there was still some disagreement: half of the group thought that knowledge, attitude and behaviour were so closely interlinked and informed by each other that they are all part of the abstract concept energy literacy; the other half advocated that literacy equals knowledge and forms the foundation for everything else.

3.3.2 How can we Measure Energy Literacy?

To address the second lead question, the researcher presented energy literacy measures on the slides for Discussion II.I. For Activity II, the researcher handed out the energy literacy questionnaires which participants read silently to assess the items in them. This was followed by Discussion II.II which gave participants the chance to discuss which questions they found suitable to measure energy literacy. Due to time issues and the dynamics in the groups, these three modules of the focus group blended into one and are presented as such, separately for academics, users and industry.

3.3.2.1 Activity II and Discussion II

3.3.2.1.1 Academics

In Discussion II.II, it turned out that in Activity II, the academics had spontaneously focused on the knowledge items only. They had disregarded the attitudinal and behavioural scales of the energy literacy questionnaire. A3 explained:

> *'I think, knowing how much energy is consumed in the home over the course of a year, or where the energy comes from, is quite useful. But I wouldn't say that therefore we should reduce emissions is a measure of energy literacy. I think it's a good thing to do obviously but I don't think it's… otherwise you're mixing normative with descriptive'.*

A1 said:

> *'It's sort of beliefs about whether you should be using oil or renewables. That is not about understanding where the energy comes from and so on. That's about whether you think it's important, whether you believe climate change is happening, whether you think it's important to do something about it'.*

The other participants in the group agreed and confirmed that they had skipped the behavioural and attitudinal items. The researcher double-checked if they would exclude attitude from energy literacy and the group confirmed. They saw the attitudinal and behavioural scales as separate questionnaires that one might wish to administer, too, but they would exclude them from the concept *energy literacy*. The remaining discussion then focused the items from Brewer's knowledge scale.

In Activity II, participants had been asked to tick the questions they thought were appropriate to measure energy literacy. The multiple-choice questions which more than half of the academics had checked were the following (with four to six ticks per question out of the six participants who handed back their printouts):

> *'Electrical power is commonly measured in units of: volts (V)/watt-hours (Wh)/joule (J)/watts (W)/British Thermal Units (BTU)/choose not to answer.'*
>
> *'What is the primary cause of current climate changes? carbon dioxide released from burning fossil fuels/there is no cause, climate change isn't real/*

natural solar cycles⁄ radioactive waste from nuclear power plants/melting glaciers in Greenland/choose not to answer'.

'Order these types of light sources from lowest to highest power usage, assuming they provide the same amount of light: incandescent bulb/compact fluorescent lightbulb (CFL)/light-emitting diode (LED)'.

'Order these appliances from lowest to highest power usage: desk lamp with compact fluorescent lightbulb (CFL)/mobile phone charger (while charging)⁄ plasma TV⁄ microwave/laptop'.

Discussing the questions, A4 said:

'I think it brings up a question which I wonder if anyone addressed through research. Which is, can you explain, can you get people to think about comparison of different appliances' consumption. I'm thinking about the questions towards the end, the one that says… number 12 [A compact fluorescent lightbulb (CFL) uses 13 W. If it is run for 2 hours, how much energy does it use?] and 13 [If your game console uses 200 W when turned on, how much energy would it waste if you left it on all weekend while you were away?]… so, 12 and 13 versus 1 [Electrical power is commonly measured in units of] and 3 [Electrical energy is commonly measured in units of]. My initial feeling was I wouldn't, we shouldn't include questions 1 and 3. But then I was thinking ok, but I definitely want questions 12 and 13. But then how do I resolve that?'.

A3 said he had the same issue. But P2 felt that questions '*12 and 13 aren't good because (…) dealing with numbers is very challenging for people*'. To that, A4 said he felt people did not need to know the exact numbers, but they need to know that '*forgetting the light on consumes less energy than forgetting your console on*'.

The researcher then presented the slide showing how studies did indeed use such comparisons to measure energy literacy. Yun et al. (2010) used a ranking task for household appliances, Anderson and White (2009) a power-rating quiz, and the ENLITEN game (www.cs.bath.ac.uk/enliten) presents two appliances and asks the player to decide which one consumes more. A4 approved verbally ('*I like that*') and the others agreed, saying that relative comparisons (such as comparing the energy consumed by lighting versus energy consumed by the washing machine) are more important for people to know and thus better questionnaire items than the numeric questions that used units and required calculations. Participants discussed that many of the items were too technical and difficult or subject to change over time (e.g. calculating the capacity of solar panels).

3.3.2.1.2 Users

The user group dismissed the questions by Brounen et al., in the discussion, judging that the questions rather asked *Do you use a lot of energy?* than *Are you literate?* and therefore the users deemed them unsuitable to measure energy literacy. Counting the questionnaire items with the most agreement was of limited meaningfulness in the focus group with the users – U1, U2 and U4 had ticked every single question, U5 had written notes in his handout but not ticked or crossed out any of the questions. The item from Mogles et al. that U3, U6 and U7 agreed on was:

> *'Which of the following actions, if everyone did this all the time, would save the most energy in the UK? turn off lights when they are not in use/turn down the heat in rooms/reduce water consumption/walk or cycle short distances instead of going by car/turn appliances off at the plug'.*

The items from Brewer (2013) that U3, U6 and U7 agreed on were the following:

> *'Order these types of light sources from lowest to highest power usage, assuming they provide the same amount of light: incandescent bulb/compact fluorescent lightbulb (CFL)/light-emitting diode (LED)'.*
>
> *'Order these appliances from lowest to highest power usage: desk lamp with compact fluorescent lightbulb (CFL)/mobile phone charger (while charging)/ plasma TV/microwave/laptop'.*
>
> *'What are the expected long-term effects of current climate changes? a significant rise in the sea level/global temperatures increasing by a few degrees on average/increasing sea water acidity/changes in seasonal rainfall patterns (droughts, floods)/all of the above/choose not to answer'.*

U5 found that Brewer's questionnaire covers scientific aspects. But U7 found:

'some of the questions are more math questions that you can answer by replacing it by anything... 13 eggs 13 watts 13 I don't know you can answer it without knowing anything about energy'.

U3 agreed and she had the concern whether

'if you're not good at math, is it not possible for you to be energy literate?'.

U2 on the other hand said:

'it's important to compute your energy consumption. If a couple of calculations are involved. Basic math. Addition, multiplication'.

U1 countered that the ordering and ranking questions were targeting the same aspect, without requiring math.

The user group did not strictly dismiss the attitude and behaviour scales (as the academics did). They considered whether attitude and behaviour could be applied energy literacy and the behaviour could reflect what they know. The discussion then drifted slightly away from the questionnaires, towards what participants knew about their energy consumption and how different billing systems (yearly bills, monthly bills, key meters) affected their knowledge. U3 summarised the discussion by saying:

'So maybe the literacy is knowing when you need to ask a question (…) Maybe it's something that we have to continue to do, continue to be 'energy literaceers', or 'energy literacering' (laughing)'.

3.3.2.1.3 Industry

In the industry group, the items from Mogles et al. that more than half of the participants agreed on were the following:

'Most of the renewable energy in the UK comes from which of the following sources? Solar/Water (hydro/tidal/wave) power/Wind/Landfill gas/Geothermal/Don't know'.

'Which of the following actions, if everyone did this all the time, would save the most energy in the UK? turn off lights when they are not in use/turn down the heat in rooms/reduce water consumption/walk or cycle short distances instead of going by car/turn appliances off at the plug'.

'Which kind of lighting uses the least amount of energy? standard light bulbs/ low energy light bulbs/fluorescent lights/LED lights/don't know'.

And the items from Brewer that more than half of the participants agreed on were the following:

'What is the primary cause of current climate changes? carbon dioxide released from burning fossil fuels/there is no cause, climate change isn't real/ natural solar cycles/ radioactive waste from nuclear power plants/melting glaciers in Greenland/choose not to answer'.

'Order these types of light sources from lowest to highest power usage, assuming they provide the same amount of light: incandescent bulb/compact fluorescent lightbulb (CFL)/light-emitting diode (LED)'.

'Order these appliances from lowest to highest power usage: desk lamp with compact fluorescent lightbulb (CFL)/mobile phone charger (while charging)/plasma TV/microwave/ laptop'.

'A compact fluorescent lightbulb (CFL) uses 13 W. If it is run for 2 hours, how much energy does it use? 13 Wh/7.5 Wh/26 Wh/130 Wh/52 Wh/choose not to answer'.

Discussing the questionnaires, the industry group reached more of a consensus than before that it was the knowledge scale that best measured energy literacy. The participants who had before included attitude and behaviour in energy literacy now changed their minds. I2 assessed that the attitudinal scale addressed environmental implications rather than energy literacy. They did consider, though, that given their perspective as a utility, householders' attitude and behaviour were ultimately what mattered because it determines their consumption. They were not sure if changing customers' energy literacy was necessary, because their goal is to provide services and products that nudge householders to be more efficient without having to learn. Again, they were returning to the train of thought that learning would still take place indirectly and hence knowledge behaviour and attitude are intrinsically interlinked.

3.4 Discussion

3.4.1 A Definition of Actionable Energy Literacy

We define energy literacy as knowledge about energy. DeWaters (2007, 2011, 2013) and Brewer (2013) define energy literacy as energy-related knowledge, attitude, and behaviour. While these three aspects are interlinked, they remain independent concepts that cannot be summarised in one term (Mogles et al., 2017). Knowledge determines attitudes to a certain degree, and knowledge and attitude determine behaviour to some extent, but attitude and behaviour are influenced by many other factors, not at least situational factors (Ajzen & Fishbein, 1977; Schrader & Lawless, 2004). Evidence of householders not wanting to negotiate practices independent of how much energy they consume have been demonstrated both in previous work (Pierce et al., 2010) and in this study.

Participants' discussion in this study confirmed that all three aspects are important in the endeavour of reducing consumption and emissions, but energy literacy should be understood in the sense of the original meaning of literacy. Literacy is defined by the Oxford English Dictionary as the *quality, condition, or state of being literate; the ability to read and write*. This definition clearly refers to an ability, i.e. the knowledge how to read and write. Equally, energy literacy is the ability to read and understand or to talk about energy matters. Whether one identifies as environmentalist (attitude) or chooses to reduce energy consumption (behaviour) are different questions. However, the goal of increasing energy literacy is to change people's behaviours. In line with social practice theory (Hargreaves, 2011), interventions aiming to increase energy literacy should focus on knowledge that is actionable, i.e. they should give feedback that householders can immediately act upon.

In line with Gabe-Thomas et al. (2016) and Hofman (1980), we define actionable energy literacy in the context of home energy use as practical knowledge. First and foremost, practical knowledge is the understanding how much energy appliances in the home use. Our definition of actionable energy literacy is very close to the examples that Brewer (2013) gives: understanding the difference between power and energy, and knowing that a microwave uses much more power than a fridge but that the fridge consumes more energy over time. Actionable energy literacy would involve householders’ understanding of how much energy practices in the home consume and the ability to critically reassess their practices and to identify opportunities to optimise consumption if they wish. This definition also follows from the focus group findings. Participants emphasised that practical knowledge (as opposed to technical knowledge), was most important, and sufficient. They thought that householders don’t have to have profound scientific understanding to be energy literate about their resource use at home.

3.4.2 Measuring Energy Literacy

In line with our definition of actionable energy literacy, measures should focus on actionable knowledge of householders. We know from previous research that people have a poor understanding of how much energy appliances consume, for example, they systematically overestimate the energy consumption of small but salient devices such as lights (Attari et al., 2010; Darby, 2006; Neustaedter et al., 2013). The hope of smart energy feedback is to correct such biases and teach householders how they are using energy for everyday practices.

To understand how much energy a device consumes, one needs to know its power and its duration of use. Unavoidably, this involves energy units (kilowatts and kilowatt-hours) and basic calculations. This involves maths skills, also referred to as *numeracy*, which has been found to be low in the general population and even in highly educated people (Attari et al., 2010; Lipkus, Samsa, & Rimer, 2001). However, exact numbers, for example whether a fridge consumes 100W or 180W, are mostly important for purchasing decisions. Given the appliances one already has, absolute numbers are less important than the general awareness how appliances (or practices) rank. Ranking tasks and quizzes have been used successfully to assess energy literacy, but to the best of our knowledge these are in studies and reports that have not been published in peer-reviewed journals (Anderson & White, 2009; Yun et al., 2010, ENLITEN energy game). Yun et al. suggest that a simple self-classification, i.e. asking people to rate on a Likert-scale how energy aware they are, is a sufficient measure to determine a person’s energy awareness.

Participants in this study deemed those items from Brewer’s (2013) and Mogles et al.’s (2017) questionnaires best that avoided energy units and calculations, and instead asked the respondent to rank appliances and practices in terms of how much power they consume. Practical knowledge would include knowing that an energy-efficient light bulb contributes far less towards the energy bill (even if left on for long periods) than most other activities in the home. While energy units are inevitable on the bill, they can be mostly avoided in questions assessing householders’ energy literacy. Participants suggested that lengthy questionnaires that include physics are unnecessary.

To assess whether householders have practical energy literacy, we need to know: Do they understand the energy units on their bill? Do they know where they are consuming the biggest share of their energy? Do they know how everyday practices and appliances compare to each other? The idea of tests in general is that one cannot only pass or fail, but reach any score in between on the spectrum and our participants emphasised that there are different levels of literacy. For diagnostic testing, this means that energy literacy tests should offer items of varying difficulty, with a few items that are very easy and a few items that are very difficult, and more items with difficulties ranging in between (assuming energy literacy is normally distributed). However, the focus of this thesis is not to develop a new test.

This thesis aims to investigate what householders learn from smart energy feedback. To that end, we will focus on householders’ knowledge about the energy they consume for everyday practices. To measure this knowledge, the following studies in this thesis will combine subjective measures (Yun et al.’s self-assessment) and objective measures (questions with correct answers), to test and compare between measures and studies which energy literacy measures work in the context of home energy use. Yun et al. found that a simple self-assessment correlated with observed energy literacy. If that is the case, it would be a quick and easy way to assess participants’ literacy without exposing them to more questionnaire items. The following studies will also include knowledge items (e.g. What is the unit that power is measured in?) to assess objectively how much participants know and if that correlates to the self-assessment. Further, to avoid confounding with numeracy, the following studies will test alternative ranking tasks where participants have to compare between common household appliances.

3.4.3 Limitations

This section discusses the limitations of the study. The user group was a convenience sample recruited at UCL and was therefore not fully representative of the general population, as all participants had a university degree. All students and staff worked on topics unrelated to energy and they did not appear to know more about energy than the average person would.

There were slight differences between groups, as outlined in the Method (3.2) and occasionally slides were skipped because of shortage of time. Another limitation was that participants did not consistently tick or cross out all items in Activity II so a consistent analysis or item count was difficult.

Finally, this study was tailored towards home energy use because this is the research focus of the researcher. It is worth noting that the energy system involves much more than just electricity and gas in the residential sector. It would certainly be interesting to investigate energy literacy with regards to transportation, consumer choices etc. and look at a holistic picture of lifestyle choices, but this is beyond the scope of this work.

3.5 Conclusion

Study 1 addressed RQ 1: What is energy literacy? The purpose of the study was to investigate existing definitions and measures of energy literacy for home energy use. Drawing from the focus group findings, we define energy literacy as *actionable knowledge about energy consumption in the household*. Behaviour and attitudes are important, but are separate from knowledge and should not be summarised under the term energy literacy. In this regard, our definition is different to the one established by DeWaters (which is used predominantly when the literature talks about energy literacy). Our definition is in line with Mogles et al. (2017) who suggest that the three concepts should be separated. When measuring what householders know about energy, we should focus on *actionable energy literacy*, such as *knowing which appliances in the home consume more energy than others, how much energy an appliance consumes over time, and how to save energy*.

Chapter 4 Interviews with Users

4.1 Introduction to the Interview Studies

This chapter addresses RQ2: How do householders interact with smart energy feedback?

While numerous studies have investigated the effect of eco-feedback on behaviour change, it is rare that they explicitly investigate data comprehension, i.e. how householders read and reflect on energy data (Fischer et al., 2013; Yun et al., 2010). It is known that householders do not engage much with their energy bills and if they do, they struggle to understand them (Kempton & Montgomery, 1982; Neustaedter et al., 2013). The main problem with bills is that they give consumers aggregated information: a single figure is reported to consumers, which describes all household energy activities over a long period of time. The question is whether smart energy feedback is more useful to householders than conventional energy bills used to be.

Current Smart Meters and most Residential Energy Feedback Systems do not disaggregate consumption data into activity-specific information, but they do provide near real time feedback. Having access to immediate information on energy consumption is a big advancement and this has been identified as a factor that encourages energy savings (Fischer, 2008). Given that disaggregation requires more complex technological solutions which haven't been optimised yet (Zeifman & Roth, 2011), this chapter focuses first on currently available feedback systems that record and feedback near real time data of total consumption.

Study 2 describes the results of interviews with householders who have SMs with IHDs. Study 3 presents interviews with householders that have been using a commercial feedback tool recording electricity consumption. Study 4 presents data from participants who received a non-commercial prototype that recorded total consumption but allowed householders to annotate the data to keep a digital diary of their household practices. These three tools were chosen to provide a probe of the different available REFS, ranging from SMs and IHDs provided by utility companies as a result of the government mandated rollout, commercial products that tech-savvy early adopters can purchase, and prototypes that are often used in research.

4.2 Introduction to Study 2

SMs with IHDs are being rolled out to households worldwide, which raises questions about whether they overcome the shortcomings of conventional energy bills. In this section, we briefly review three studies that have investigated how householders interact with smart energy monitors to derive the research questions for Study 2.

Hargreaves et al. (2010) presented one of the first attempts to understand how UK householders interact with smart energy monitors. In a qualitative study, they provided 15 households with smart energy monitors and interviewed them about their motivation for participating in the study, how they used the monitor, and whether the monitor had changed their awareness or behaviour. Participants were motivated by an interest in technology and information gathering, and to varying degrees were interested in reducing their consumption for either financial or environmental reasons. During the study, when the monitor showed consumption increasing, participants would react by going around the house and switching things off. Occasionally they would also delay practices, for example, they would decide to run the washing machine the next day. These reactions were mostly novelty effects that wore off over time. A central observation in this study was that first, the smart energy monitor needed to be placed in a location where it could be seen and second, it was therefore important that the monitor was aesthetically pleasing.

In a similar study, Wallenborn et al. (2011) installed smart monitors in Belgian households. The aim of this study was to understand what users learn from these monitors, and whether they experience a change of perception and behaviour. While Hargreaves et al.'s sample consisted of tech-interested early adopters, Wallenborn et al. took care to recruit a more representative sample, including low-income households. The study confirmed most of Hargreaves et al.'s results, but found that smart energy monitors can only change behaviour in users who were already interested and involved in saving energy before participating in the study. Overall, they found that current smart electricity displays were poorly designed for most users.

Strengers (2011b) found across three trials in Australia that IHDs have benefits, but they also bear risks. Strengers found that householders were often alerted to increases in their usage by spikes in graphs of energy consumption or flashing lights emitted from the IHD. Noticing these spikes led to behaviour change in some cases. The problem of legitimising practices mostly arises from lights with traffic light systems, where green light indicates low consumption, amber indicates medium, and red indicates high consumption. The green light is interpreted as *good,* amber as *ok*, and red as *bad.* If the light remained amber and did not change to red, participants felt reassured about energy-intensive practices such as using the tumble dryer and continued their habits, feeling they were legitimate. They also neglected practices that are considered non-negotiable for cleanliness, comfort, and convenience reasons (Shove, 2003). The colour-coding disguised the impact of practices and thus they were overlooked, which is exactly the opposite of what smart energy feedback is hoped to achieve. Ideally, users would reflect on everyday practices and try to identify what they could change to reduce their consumption. Strengers also points out that if the design of smart feedback is producer-led, there is a risk of them being most interested in offering targeted services, rather than the design process being user-centred. Indeed, the rollout of SMs and IHDs is mandated by the UK government, they are provided to householder by utility companies and manufactured by third parties.

In summary, the previous publications leave several questions unanswered. Firstly, studies aim to understand what people learn and whether they change their behaviour. They do not necessarily focus on understanding *how* users learn, i.e. none of the studies reports detailed contextual inquiry data to investigate how participants read the displays and reflect on the information in situ. Strengers (2011b) reports that participants did *not* reflect on practices and neither did they make changes. Both theory on persuasive technology (Li et al., 2010) and work with a focus on practical interventions (Ploderer et al., 2014) have identified reflection as a crucial step for users to gain insights and change their behaviour. Second, most studies recruit samples and provide them with smart monitors. Hargreaves, Nye, and Burgess (2010) interview a sample of early-adopters, and while Wallenborn, Orsini, and Vanhaverbeke (2011) recruited a more balanced sample, they still equipped participants with electricity monitors for the sake of the study. SMs and IHDs on the other hand are rolled out to UK households without users' choice to opt-in. Third, Hargreaves et al. found the most promising results, but their monitors gave disaggregated feedback on the appliance level. SM IHDs currently do not provide this feature. Fourth, an aesthetic design of the monitor and its display turned out to be central. The SMETS merely specify that the IHD should be clear and easy to understand, but there are no design criteria with regards to their aesthetics (DECC, 2013).

The purpose of Study 2 is to examine how UK householders who already have a SM IHD read and reflect on the information display. To this end, we recruited participants that had been given a SM with an IHD by their energy providers as part of the UK rollout. We interviewed participants and where possible we met them in their home to conduct a contextual inquiry in which their task was to walk us through the information displayed on the IHD.

To assess how energy literate participants were, we asked them a few questions before the interview that are in line with the results and discussion of Study 1. According to Yun et al. (2010), it is sufficient to ask users to rank their literacy on a Likert-scale. To read energy information customers need to be able to deal with energy units such as kilowatt-hours, so we added three questions requiring units and a basic calculation (Brewer, 2013). Seeing that energy units are very technical, we added a more pragmatic question to see if participants could intuitively rank their household appliances in terms of how much energy they consume (Yun et al., 2010; Anderson & White, 2009).

We then conducted a contextual inquiry (Holtzblatt & Jones, 1993) using the think-out-loud method (Lewis & Rieman, 1993), i.e. householders talked us through the information displayed on their IHDs. All interviews were audio recorded and transcribed in the transcription software f5. The transcripts were imported into the qualitative data analysis software Nvivo and analysed thematically (Aronson, 1995; Clarke & Braun, 2014; Seidman, 2013).

4.3 Method

4.3.1 Sample

Six participants (3 female) were recruited because they had a SM in their home. Participants were recruited through word of mouth and posts on social media. The average age of the sample was M = 31.6 years (SD = 3.1) All participants had a university degree. One participant lived alone, four lived with their partner or one flat mate, one participant lived with two flat mates. The households' energy providers were British Gas (2x), E.ON (2x), Co-op and Bulb (P5 had received her smart meter from EDF but then switched to Bulb). At the time of the interviews, participants had their Smart Meters for between 1.5 months and 1.5 years. They said they spent between £20 and £40 on electricity per month, £70 and £75 in two cases were gas was included in the bill.

4.3.2 Materials

Participants were recruited because they had a Smart Meter in their home. They were not provided with further equipment. The interviews were unstructured because the context was different for every participant.

Demographics were collected during the interview. The demographics questions included five questions to assess participants' energy literacy. These were a mix of energy literacy questions discussed in Study 1:

1. *On a scale from 1-5, how much would you say you know about your energy use at home?*
2. *Electrical power is commonly measured in units of…?*
3. *Electrical energy is commonly measured in units of…?*
4. *A lightbulb (CFL) uses 13 W. If it is run for 2 hours, how much energy does it use?*

5. *Please pick five appliances in your household and rank them in terms of how much electricity you think they consume and explain why.*

4.3.3 Procedure

P1, P2 and P3 took part in a contextual inquiry about their IHDs. P4's IHD was not working and P5 and P6 had disposed of their IHDs, so contextual inquiries were not possible; they were interviewed on their experience with their SM and IHD. P1, P2 and P4 were interviewed in their homes. P3 was interviewed via Skype upon her request. P5 and P6 were interviewed in a public space. Interviews lasted from 15 to 25 minutes, with an average duration of 18 minutes (SD = 3.5).

4.4 Results

4.4.1 A Priori Energy Literacy

In the self-assessment of their energy literacy, participants scored on average M = 3.4 (SD = 0.75) on a scale from 1-5. P3 answered two out of three questions on power and energy correctly; P2, P4, P5, P6 and P7 answered none correctly. Participants were asked to pick five appliances in their household and to rank them in terms of how much electricity they consume. All lists were assessed as reasonable (e.g. P6: tumble dryer, underfloor heating, dish washer, TV, lighting, in descending order).

4.4.2 Interviews

The qualitative data from the interviews is presented in two parts: first, the contextual inquiries with the three participants who were using their IHD; second, the interviews with the three participants who were not using their IHD. User requirements are presented that emerged both from the data of both groups (IHD users and IHD abandoners).

4.4.2.1 Contextual Inquiry

We first report the themes from the contextual inquiries with the three households whose IHDs were up and running.

4.4.2.1.1 Location

Where do householders place their IHDs? P1 and P2 had their IHD sitting on a kitchen shelf. P2, who had her smart meter for 1.5 months, kept the IHD at eye-height where it was well visible. P1's IHD sat on a higher shelf where it is '*out of the way*', meaning that only P1 herself could properly see it, whereas her two flat mates could not because they are shorter. P3 had hers on the floor in the main area of her small studio flat, which was determined by the fact that is must be constantly plugged in to be powered and all her sockets are near the ground.

4.4.2.1.2 Default Information

The IHDs' default screens show the energy consumed so far – in the case of P1 and P3 it shows the energy consumed so far on the day, for P2 it shows the energy consumed so far for the running month (Figure 2). P1 usually looks at this figure '*in the mornings when I'm kind of waiting for the coffee or something like that and then its normally around about a Pound*'. She uses it to contrast every day and to see if it is higher than it normally is. P2 sees the money spent for the month so far: she was interviewed on the 2nd of the month and explained that their energy had cost her 2.72 Pounds in the last two days. She said the day before it was 67 Pence in the morning right after getting up and that made her reflect how she used those 67P ('*What have I done? Already?')* and she concluded it must have been the shower and the kettle. When going through the menu in the contextual inquiry, she found that she'd spent 1.19 Pounds so far that day ('*Wow, I didn't know I could see this!*') and explained the difference between the day before and that day by saying: '*I guess I am home*'.

Figure 2. P2's IHD by E.ON.

4.4.2.1.3 Traffic Light Colours

IHDs come with coloured LEDs at the bottom of the IHD in green, amber and red show if consumption is low, medium or high. The British Gas IHD also comes with a colour display which shows a gage in the same traffic light coding (Figure 3). The picture taken of P1's IHD shows only one green bar, indicating that the consumption is very low. Figure 4 shows an example of the colour-coded gauge going up.

Figure 3. P1's IHD by British Gas.

Figure 4. IHD by British Gas, example of the gauge.

P3 used the colours intuitively when flicking through the dates: '*Oh god, yesterday was cold'*, inferring that she used a lot of energy for heating as indicated by the red colour for that day in the calendar view (her display shows electricity use only because she does not have gas and everything in her flat runs on electricity). She mentioned that in the summer, '*it's kind of harder to tell',* because she does not heat, so she only checks the IHD once or twice every few weeks. In winter, ('*when I think electricity costs become a problem'*), she said she uses the IHD daily and when the light goes amber and red she would try to cut down her use, i.e. she would try to heat less and shower and dry her hair at the gym. She said she still needs to look at her bill to find out how much she is spending over time, but the IHD helps with her daily budgeting:

> *' it's been quite good because I can see every day what I'm using (…) when I was feeling like oh my god I can't afford it I just switched off everything (…) it makes you realise what you can do to save energy. Cause sometimes when you just pay a bill at the end of three or four months you don't realise what you're using. So, this is quite a good way to keep track of what you're using. For example, I looked at it yesterday and I saw it was nearing the oranges and red, then I know it was probably a very cold day or it was a day off when I spent quite a bit of time at home and had to heat (…) helps with my internal budgeting in my head'.*

P1 referred to the colour-coded bars (Figure 4) as the '*wheel of fun'* and explained:

> *'When we use electricity it goes up and roundabout here it turns colour into yellow and red (…) The guy who installed it said you obviously want to keep that very low always and I thought 'No, I want to see how high it can go up' (laughing). So, I find it really interesting to see what kinds of devices use peak electricity. So, the kettle is obviously quite good. And the washing machine sometimes. That goes into the red. Everything else is a bit boring to watch'.*

4.4.2.1.4 Useful Functionality

Participants briefly talked about other functions of their IHDs that they use less frequently. P3 demonstrated how she can see weekly, monthly and yearly consumption beyond daily use and explained she can also get information about her billing cycle. She found it helpful that the IHD gives out her meter reading which she needed when changing provider and that prevented her from having to climb up to the meter. P1 used the IHD to find out how much money they have left in their prepaid electricity account to check when it is time to top up. P2 demonstrated in the interview how to find the tariff info, which tells her how much the kilowatt-hour is currently costing her depending on the time of day as she is on a time-of-use tariff. She reported that in the mornings the IHD tells her how long they have until the tariff's changing from the cheaper night rate to the day rate.

4.4.2.1.5 Useless Functionality

P2 mentioned there was a budget function but she wasn't using it. She had played with other buttons on the IHD but wasn't entirely sure what they do because she hadn't read the manual. She came across the button to switch units between Pounds, kilowatt-hours and CO_2 in Kilos, but said she did not know what the CO_2 means. P2 mainly appreciated the IHD for its non-energy related information. She liked that it tells her the temperature in the kitchen and she stated that she mainly used her IHD as a clock since the kitchen clock had been out of battery for a while.

4.4.2.1.6 Behaviour Change

P3 was the only participant who reported behaviour change (heating less and showering at the gym when the colours on her IHD turned amber and red). She used the IHD to save money ('*when I was feeling like oh my god I can't afford it I just switched off everything*'). P1 said she had not changed anything since having the IHD but pointed out that she felt she had already been responsible before, as in not overfilling the kettle for example or washing only full loads of laundry. She said she was '*not gonna stop using the washing machine because of the electricity (laughing)'* and she was not prepared to buy more efficient appliances because she lives in a rented furbished flat. With regards to the feedback from the IHD, she said '*it hasn't given me any information which I could act upon'*. P2, when asked if she's gained an insight as to which the biggest contributors to her bill are, responded: '*not really'*. Her suggestion how she might be able to save 10% of her consumption was '*unplugging things'*.

4.4.2.1.7 Satisfaction

In terms of satisfaction and acceptance, participants' attitudes varied, particularly between P3 and P1. P3, who was actively using her IHD to monitor consumption and who adapted her behaviour, explicitly said she was '*happy'* with her IHD. In contrast, P1 reported that one of her housemates '*hates that there is a display that's always on'* despite her flat mate not being particularly concerned about their energy use. P1 suggested her house mates' disapproval had to do with the fact that the SM and IHD were '*this thing that came and was brought into our house'* by the decision of the utility company. P2 was happy with her IHD and thought SMs with IHDs are a '*good idea'* but she worried that people like her mum would restrict themselves too much ('*she will not turn the heating on even when she's cold because she can see the number going up'*).

4.4.2.2 Abandoned IHDs

P4, P5 and P6 were not using their IHDs. The reasons for that are explained by the following themes.

4.4.2.2.1 Damage

P4's IHD sat in the flat's storeroom where it has been since they got it. He thinks it used to work in the beginning, but at the time of the interview the screen did not work and couldn't be turned on; he couldn't say if the IHD was out of battery or broken (Figure 5). When asked if he ever engaged with the IHD when it was working, he said: '*not really if I'm being honest'.* P4 went through his bills during the interview and explains how they were charged well over £100 in June 2017 which must be inaccurate:

> *'I was only here for two weeks and my flat mate wasn't here at all. If anything, that should be without doubt our lowest month (…) when we realised it was inaccurate, that it wasn't giving decent readings at all, of our energy use, we just thought well we don't use it anymore'.*

P4 commented on his bill that he did not '*know if this balance is what we owe or what we're indebted*'. He elaborates that his flat mate got in touch with the utility but '*they really just don't care. You feel really helpless. The money was just taken out of our account*'.

Figure 5. P4's IHD by E.ON.

4.4.2.2.2 Ugliness

P5 remembers that they first put the IHD up in the kitchen but there was no free plug, so they moved it to the dining room where it was sitting on the floor. He said he '*was gonna mount it'* to the wall but never did and eventually the IHD went missing when they redecorated the house. He said the only time he '*was interacting with it was when it was in the kitchen when it was new. And that was probably no more than a week, maybe two weeks'*. The main reason for abandoning the IHD was, other than the inconvenience of having to plug it in and mount it, that it was ugly. P5 said his IHD was:

'like the old CRT TV screens that you know … you needed a bigger living room in order to watch TV (laughing). It has that massive back and a small rubbish screen (…) it wasn't on all the time. You had to go up to it and you had to press a button and then it took a couple of seconds and then it would come on (…) it is really, really ugly. Because it's that old CRT type looking unit. Like it would stick out a lot. And also requires to be plugged in. So there would be a cable running down to a plug. And it also limits where you can mount it on the wall because you have to mount it close to a plug (…) It was like looking at a mobile phone from 10 years ago. It had some really basic colour but I mean it was a really pixelated interface. It was just not very nice to look at. When you were changing screens, it would have that fade in fade out that you'd get from these old LCD watches you know'.

4.4.2.2.3 Futility

P6 disposed of his IHD. He used to have it in the living room next to the speakers, and he said '*that area felt cluttered'* and '*looks a lot nicer now'* since they threw it out. P6's reason for disposing of his IHD was that he perceived it as useless:

'It was taking up space. Using electricity (laughing)'.

While they had it, he said he looked at it once or twice, but did not end up using it and he summarised:

'it just felt like another device that was plugged in, always on, and for the little amount of time that we used it, as we never looked at it, it felt excessive (…) we were just doing a big clean up, and as we had all of these old devices, all of these old things that need to get recycled, so we went 'this is something we never use, grab it, and (…) Completely threw it out (…) the camera [Nest camera] actually found somebody going through our trash and take it (laughing). So somebody else out there now has my British Gas head unit (laughing). So, I hope it brings them more joy than me (laughing)'.

4.4.2.2.4 Useful Functionality

Like P3, P6 said the real benefit of the SM was the automated meter readings being transferred to the utility:

> *'before they installed the smart meter, I always, every three month or so, I had to take a photo or enter the meter reading into the website and I don't have to do that anymore so that's great! That's perfect!'.*

He later added:

> *'For the most part, I feel like the 'smart home' today is just a remote control experience and it's not actually smart'.*

4.4.2.3 User Requirements

Across all participants, themes emerged that can be summarised under user requirements: smarter feedback, definitions of goals and values, and usability improvements and gamification.

4.4.2.3.1 Smarter Feedback

Participants asked for *smarter* feedback, including actionable appliance-specific information, baseline references and social comparisons. P4's IHD was not working but he would have liked to see his consumption data to work out which appliances use a lot. He disliked the idea of waste for ecological reasons: '*the main thing would be that it's better for the environment. The cost savings would be pretty marginal*'. He further said:

> *'it was (…) like 'Oh you can then see when your energy use spikes' but it didn't actually tell you, like, what device is using a lot of energy. It's like yeah I use my electric kettle in the morning, I expect my energy to spike (laughing), because I'm boiling water'.*

The only participant who was satisfied with the feedback and who derived implications for behavioural change (i.e. turning the heating off, showering at the gym) was P3. P5 and P6 compared the IHD to their smart thermostats, which they are happier with than with the SM. P5 and P6 thought it would be of value if the IHD offered consumption information on long-term trends, references and comparisons like the smart thermostats do. P5 said his smart thermostat provides statistics about how much energy or money have been saved over the last weeks or months. P6 explained that the comparison functions of the Nest thermostat are useful and he would like baseline references and social comparisons for his electricity use, too.

4.4.2.3.2 Definition of the IHD's Goals and Values

It was unclear to participants what the value and goal of the IHD are.

> *'Does it help me save money? Does it help me protect the environment? The value wasn't articulated at all (…) I guess this is where my biggest question was. I don't know what they want me to change and if I would actually even change it. I'm gonna make my coffee before I go to work (…) I'm gonna watch TV when I get home from work, you know. Maybe some things like… putting on the washing machine at a different time (…) for me the value of the smart meter was the actual meter itself and not having to add another screen to my life. That's where I feel like less screens are better' (P6).*

He further suggested the feedback could be more beneficial in an entirely different format and without an IHD. Instead, it could be delivered via app or his utility could send him a monthly email report. P1's flatmate, too, disliked that there was another display that's always on.

4.4.2.3.3 Usability Improvements and Gamification

P5 on the one hand found her IHD intuitive to handle without ever having looked at the manual. On the other hand, P1 and P2 found the IHDs' controls less intuitive. P2 reported:

> *'I've played around with the buttons, I'm not entirely sure what they do. I dropped it one time and the display went off and I tried to get it back (laughing). But I didn't really know and that's because I haven't read the information leaflet'.*

During the contextual inquiry, P1 sometimes did not know what she was looking at:

> *'I have to admit I don't know if this is the daily use… no this is the use I am using at the moment so within this hour? Sorry I don't know how I got to this display. And I don't know how to get back to the start screen (…) Sometime in the evening the display is off, it goes into some kind of stand-by mode. I don't know how to make it or how to get it out of it but that's that (…) It has a lot of functions that we don't use'.*

P1 was also the participant who used the traffic light gauge in an unintended way, instead of keeping it in the green area, she would try to see how high she can drive it up. She suggested to gamify the feedback with a '*visual reward to get something out of keeping it as low as possible'*. P6 mentioned he was fond of his Nest thermostat and appreciated its gamification feature.

4.5 Discussion

4.5.1 Main Findings

The interview with P4 confirmed that he found his conventional energy bill confusing and not useful (Neustaedter et al., 2013). The question was whether smart IHDs are more beneficial. This study found that more than half of those interviewed with a SM were not interacting with their IHD at all. One IHD was broken, one was lost, one intentionally disposed of. The finding that people abandon smart devices is consistent with previous research (Lazar, Koehler, Tanenbaum, & Nguyen, 2015), and Hargreaves, Nye, and Burgess (2010) describe that householders quickly neglect smart energy monitors as soon as the novelty effect of owning a new device wears off. In comparison, a report by DECC, including 2,000 Smart Meter users, found that 'six in ten (61%) reported that they generally still had their IHD plugged in, while two in five (39%) did not' (DECC, 2015, p. 42). The report further found that only 4% were dissatisfied with their Smart Meter including the IHD. However, engagement with and understanding of the IHD varied greatly between participants, and it turned out that some who said the IHD was 'easy to use' were only using the traffic light function (DECC, 2015, p.7), just like one of our participants. In-depth interviews revealed that even participants who were engaging with the IHD did not necessarily understand how to interpret it or how to make behavioural changes.

4.5.2 Reflection

The aim of this study was to observe how householders read and reflect on the information provided by the IHD. We managed to interview only three participants in a contextual inquiry, asking them to walk us through the information in the IHD. The default screen of these IHDs reflected changes in consumption in near real-time, with a numeric figure indicating how much energy had been consumed so far. In theory, the immediate feedback holds potential for householders to reflect-in-action and learn how much energy they are consuming while carrying out a practice (Schön, 1987). However, during the interview, participants did not link the real-time information on the display to their energy use in that very moment. They reported that since they received the IHD, they had identified appliances that made the figure or the traffic lights spike up, indicating that some reflection-in-action had taken place with the help of the IHD.

Albeit, the reflection was limited to a couple of insights per person (i.e. identifying the kettle and the washing machine as appliances with high wattage). Only one participant reported adapting her behaviour with the help of the ambient feedback light system, actively monitoring her consumption and keeping her heating in check in the winter and showering at the gym. Firstly, the latter demonstrates that people do not necessarily reduce their consumption, but merely shift it away from the home. Secondly, this participant said she knew before that heat inducing devices use most energy. Other than confirming that heating and showering use a lot of electricity, she did not gain any new insights from interacting with the IHD.

Hargreaves et al. (2010) reported that users would show signs of reflection-in-action (Schön, 1987), going around the house and switching things off when they saw consumption rise, and reflection-on-action, changing up their routine. Behaviour change was not widespread in Hargreaves et al.'s sample, and it was even more rare in our study, as we found only one in six participants used the IHD to reduce her consumption. Other participants emphasised how they needed to continue their practices (e.g. washing laundry) or how they were not willing to give them up (e.g. having coffee in the morning). Both aspects are in line with previous work (Shove, 2003; Strengers, 2011b).

Going through the menu, it became clear that participants did not use or understand all of the provided function which also is in line with one previous study we are aware of (Yang et al., 2014). Overall, the IHD served mostly as an ambient reminder, and the rich data recorded by the smart meter was not made accessible in a way that allowed users to engage with it. Participants' interaction with the IHD in the contextual inquiry was so brief that it was difficult to observe how they read and reflected on the provided information and how they mapped it to their everyday practices.

We observed factors that restricted the usefulness of the IHD to householders. The location where householders place their IHD was found to be relevant in previous work (Hargreaves et al. 2010). However, the IHD's positioning is restricted by the requirement to be close to a power outlet. In one case, this seemed to play a role in one of our households abandoning the IHD eventually because they could not find a convenient spot for it. Even one of the participants who used her IHD mentioned that it was on a kitchen shelf out of sight for her flat mates. Hargreaves et al. found that this *out of sight out of mind* practice limited users' interaction with the energy monitors.

Energy feedback needs to be presented in an appealing way to be effective (Fischer, 2008). The aesthetics of the IHD played a major role for our participants, as was also found before by Hargreaves et al. (2010). The IHDs we encountered varied widely in their aesthetic appeal (Figure 3 - Figure 5). One participant elaborated at length how *ugly* the IHD was which was why he did not want to mount it onto the wall or engage with it. Independent of its aesthetics, the IHD was simply not perceived as adding any benefit to householders' lives.

It appears IHDs are not living up to the promise of their *smart* label (Mogles et al., 2017), and some of our participants thought of them as just another screen cluttering their homes and using electricity ('just another gadget', Hargreaves, Nye, & Burgess, 2010, p.6117). Previous work has addressed how digital technologies in the home increase energy consumption and how hardware and software - and whether they are used long-term or abandoned quickly - are part of sustainability (Davies et al., 2017). The findings demonstrate the negative effect of poor aesthetics and lack of perceived usefulness on the acceptance and adoption of technologies (Davis, Bagozzi, & Warshaw, 1989). One interviewee also brought up the perceived loss of control, referring to the IHD as something that was brought into the home by the energy provider. Ironically, smart energy feedback is supposed to give householders more control over their consumption. Yet one participant thought the benefit of the current system is merely that they communicate data back to the utility company, automatically bypassing the customer.

4.5.3 Limitations

4.5.3.1 Energy Literacy Measures

We asked participants to self-assess their energy literacy on a Likert-scale. Yun et al. (2010) found this was a sufficient measure. At this stage, it is hard to evaluate whether the self-assessment correlates with the responses to the other questions. Most participants could not answer the technical questions, but all gave reasonable answers to the ranking task. This might indicate that the self-assessment worked, as participants scored on average 3.4 on a scale from 1-5, indicating medium energy literacy. The fact that most of the sample was unable to say what the unit of measurement is for both power and energy (let alone to calculate one from the other given duration information), shows that these items can only be answered by people with very high energy literacy, but they are too difficult for someone of average energy literacy. The ranking task (naming and sorting five appliances in terms of how much they consume) sits at the other end of the spectrum and everyone gave a reasonable response. Maybe the question was too easy because we let participants choose the devices themselves. The sample was too small to properly evaluate the suitability of the energy literacy questions we asked.

4.5.3.2 Sample Size

The sample was very small. Even though qualitative studies often work with small samples (the studies reviewed in the introduction to this study ranged between 15-28 participants), our sample was only half of what would be desirable. Recruiting householders with SMs proved difficult, because it is still a minority of the population who is equipped with one.

4.5.3.3 Methodology

The aim of the study was to investigate how people read and reflect on information shown on IHDs. This was quite difficult because the default screen of the IHD contains very little information and participants did not link the figures to what was happening in the home during the interview. None of the participants had any bigger appliances running at the time, so the consumption data barely changed over the course of the short interview. The contextual inquiry was useful for participants to talk through all the menu functions and to demonstrate which ones they were using and which ones they did not use or understand. The research question how users reflect on the information and how they link it to their actions could not be addressed as thoroughly as intended, which was mostly due to the scarce data provided by the IHD.

4.6 Conclusion

Smart feedback provides more potential than conventional energy bills for users to learn about their consumption. Yet, one half of the participants in this study were not engaging with their IHDs at all. Overall, engagement with the IHD was largely restricted to using the ambient traffic lights or to checking the spending on energy in the morning to see whether the consumption was typical, or lower or higher than usual (but this was not followed by insights or actions). It seems that the information provided by the IHD is very minimalistic and do not display the rich data collected by SMs in an accessible way. The biggest constraint is that one only sees a snapshot for the very moment when looking at the IHD. The SMETS require IHDs to show real time feedback and historic data (DECC, 2013). The only data we found our participants interacting with was the constantly updating (but cumulative) real time data, but neither could participants see the wattage of their consumption in each moment, nor could they examine their history of use or trends over time.

4.7 Introduction to Study 3

Study 2 investigated householders' interaction with SM IHDs. A major constraint with the IHDs in Study 2 was that they only provided snapshots of energy consumption, which limited both the extent to which participants reflected on the data as well as the observations we could make in the contextual inquiries. Many studies on energy visualisation focus on instantaneous feedback and when they present historical data, they show a cumulative value for the consumption over the last day or week (Costanza et al., 2012). SM IHDs are not the only smart energy feedback tools, but there are plenty other REFS available and some of them visualise rich records of electricity usage data.

Study 3 describes a study in which householders were given a commercial feedback tool that displays a detailed history of electricity use via a web portal and mobile app instead of using an IHD. Instead of presenting a single value for the consumption over the past day (or week, or month) like many REFS do, we chose a tool that visualises consumption data in graphical format to show the user how much energy they have been using at any point in time (Figure 8).

Energy data is time series data and it is most commonly visualised in a line graph showing power over time (Costanza et al., 2012). This visualisation is conceptually very close to what it represents and therefore seems a reasonable way to display energy data (Pinker, 1990). The question is whether householders can successfully engage with energy data visualisations, given that the average person is neither energy literate (Attari et al., 2010; Darby, 2006) nor trained in reading graphs. Even simple bar or line graphs may be difficult to understand for many people (Galesic & Garcia-Retamero, 2011).

Graphical perception and literacy are defined as the visual decoding of information encoded on graphs, and the decoding process may require considerable cognitive effort (Cleveland & McGill, 1984). According to Murugesan et al. (2014), the visualisation of energy consumption is widely considered an important means to assist end-users in reducing energy consumption and bringing about sustainable behaviour. However, there are no clear design requirements to develop energy visualisations and if they are not chosen wisely, they can negatively affect people's data comprehension (Baur et al., 2012; Tong, Gromala, Bartram, Rajabiyazdi, & Carpendale, 2015; Tufte, 1983).

Only a few studies have focused on displaying rich time series data showing the ups and downs in power over time (instead of snapshots of instant consumption) (Broms, Ehrnberger, Ilstedt Hejlm, & Bång, 2009; Costanza et al., 2012). Broms et al. use an ambient and artistic display, the Energy AWARE clock, which is in the shape of a house, with a colour display showing electricity consumption instantly and over time in circular pattern (Figure 6). This is an abstract visualisation without an accurate power scale. Costanza et al. (2012) built a software that displays accurate power data over time in a conventional coordinate system (Figure 7).

Figure 6. The Energy AWARE Clock by Broms et al. (2009).

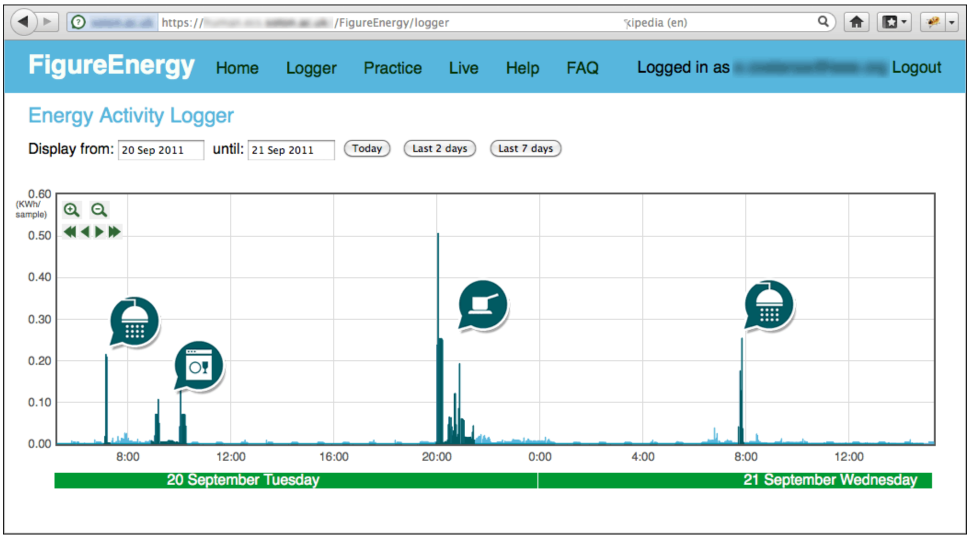


Figure 7. FigureEnergy by Costanza et al. (2012).

Designing effective visualisations for energy feedback is a challenge and modern visualisations (like the Energy AWARE clock by Broms et al., 2009) attract people but may require initial learning to understand (Murugesan et al., 2015). Therefore, we opt to investigate a traditional graph that is more appropriate for effective decision making, like the time series graph in FigureEnergy (Costanza et al., 2012).

In Study 3, we gave 13 participants from nine UK households an electricity power clamp meter. The study consists of four parts. First, to assess participants' energy literacy, we asked people which appliances in their home they think consume most energy and whether they could quantify that in any way. In addition to the verbal questions, we also asked them to make a sketch of their electricity consumption over a day. Sketching is a method typically used in HCI to inform interface design (Buxton, 2010). Like Chisik (2011), we chose this approach within a user study because sketches are a rapid, accessible and expressive method which reveals the mental model of the subject: visuals and thinking are closely linked and sketches support the thinking process because they are an externalisation of internal thought (Greenberg, Carpendale, Marquardt, & Buxton, 2011; Kirsh, 2010; Tversky, Corter, Nickerson, Zahner, & Rho, 2008; Walny, Huron, & Carpendale, 2015).

Second, we conducted a contextual inquiry (Holtzblatt & Jones, 1993) using the think-out-loud method (Lewis & Rieman, 1993), asking participants to explain the web-based time series line graph visualisation of their electricity data. Asking users to think-out loud grants insights into what they are thinking, what questions come up as they explore the data, and how they read and reflect on what they see. The semi-structured method offers the required flexibility to react to individual approaches.

Third, in follow-up interviews three months later, we assessed whether participants' understanding about their consumption had changed and whether they were still engaging with the tool or whether they had abandoned it (Lazar et al., 2015). Fourth, we wanted to derive user requirements based on participants' input. We invited participants to generate ideas how smart electricity feedback could be optimised.

In Study 3, we aimed to recruit more than one person per household. Previous research found that collecting data from more than one household member improves the results (Yang et al., 2015). All interviews were audio recorded and transcribed in the transcription software f5. The transcripts were imported into the qualitative data analysis software Nvivo and analysed thematically (Aronson, 1995; Clarke & Braun, 2014; Seidman, 2013).

4.8 Method

4.8.1 Sample

The study was advertised by posts on Twitter and Facebook, by leaflets in the university building and by word of mouth, aiming for a sample that would cover a variety of housing and occupant types. The advertisement asked if the reader wanted to learn more about their domestic energy consumption, and that they would get to keep the energy monitoring device after the end of the study as a reward for taking part. We recruited 13 participants (6 female) from nine households. Eleven out of 13 participants filled in our demographic online questionnaire. Mean age of the sample was M = 40 years (SD = 15, range 25-76). Education varied from less than high school to doctoral degree. Five of the participants lived in terraced houses, three in apartments, and two in semi-detached houses. Most of the participants lived in rented accommodation (10 of the 13 participants). None of the subjects lived alone. Two households had used a smart meter before in previous residences but none had a smart meter in their current property. None of the households were on a time of use tariff (where electricity costs different prices at different times of the day).

4.8.2 Materials

We provided each household in the study with an electricity-monitoring device: The Loop energy saving kit (available from UK-based technology company Navetas, see www.loopenergysaver.com). The Loop energy saving kit consists of three hardware items. First, there is a current clamp transmitter that measures the household's electricity consumption. The transmitter must be clipped to an electrical conductor (i.e., one of the electricity meter's cables). It transmits measurements to a receiver by radio transmission. Second, the receiver, which is plugged into the household's Internet router, communicates the household's electricity consumption data back to the server so the occupant can see the collected information online. Third, a power plug that powers the receiver. The customer is guided through the installation process on the your-loop.com website.

A key feature of the Loop energy saving kit for the purposes of this study is that it has a web portal that allows users to look at their household's electricity consumption (available at: www.your-loop.com, see Figure 8). Using the web portal, householders can see their recorded electricity usage graph on a daily, weekly, or monthly basis, in a range of unit options (financial cost as estimate in Pound sterling, energy use as estimate in kilowatt-hours, or environmental impact as estimate in CO_2 emission). We chose the Loop for our study because it is representative of the tools currently available on the market. It is very affordable and easy to install, and it visualises domestic electricity consumption as time series data.

4.8.3 Procedure

We sent the Loop energy saving kit to the participating households, including the instruction to set the device up but not log in to the website yet. All participants later confirmed that they had complied with this instruction. Two weeks after the participants had installed the Loop, we conducted the first of two interviews. The second follow-up interview took place three months later. Interviews took place partly face-to-face, partly over Skype. In the interviews that were conducted on Skype, participants would share their screen with the interviewer so both were able to see the website. Three participants choose not to take part in the follow-up interview, reducing the sample size to ten. Both interviews consisted of two parts. The procedure outlines the four parts in the following.

4.8.3.1 Interview I, Part I

In the first part of the first interview, we asked participants the following three open-ended questions to assess their energy literacy.

1. *Which electrical devices in your household do you believe consume most electricity?*
2. *In a metric of your choice, can you please estimate how much electricity your household appliances consume?*

3. *How do you think your electricity consumption look over a day? Please, make a sketch of your electricity consumption over a day.*

Participants were provided with pen and paper for the third questions. Participants were free to choose the type of graphic and metrics they wanted, and we stressed that drawing skills were not important.

4.8.3.2 Interview I, Part II

In the second part of the first interview, we would ask participants to log in to their Loop account. The task in the contextual inquiry was to verbalise what information they saw in the graph (Figure 8) and to explain which appliances or activities have led to the displayed patterns. Our semi-structured interview guide contained questions that we used to nudge participants if they stopped thinking out loud (e.g. What do you see? Can you please interpret the graph you see? Can you identify what you did? Can you identify appliances in the graph?). Interview I lasted from 25 to 45 minutes, with an average duration of 35 minutes (SD = 8).

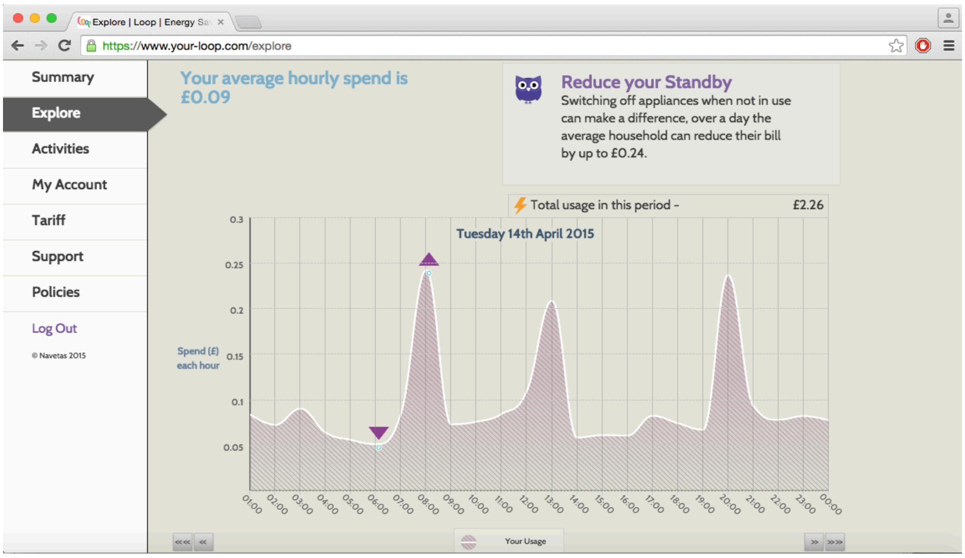


Figure 8. The energy consumption history of the Loop.

4.8.3.3 Interview II, Part I

In the first part of the second interview three months later, we asked participants whether they had logged in to the Loop website again since the first interview (or whether they had used the mobile app). If they said they had not, we followed up to inquire why they hadn't and what could have motivated them to look again. If they had continued to use the feedback tool, we inquired how regularly they had used it and to describe the situations when they looked at it. We explored their motivation for engaging with the feedback and whether they had reflected on their use and if they had learned from the Loop. If participants stated that they had learned something, we inquired what they had learned and how they learned it. If they said they had not been able to understand the feedback, we explored what prevented them from learning from the data display.

4.8.3.4 Interview II, Part II

In the second part of the second interview, we instructed participants to imagine that they were involved in designing the 'perfect' energy feedback. We stressed that they were welcome to use their imagination without considering technical feasibility. We asked them what their smart feedback would be like, which functions they thought were important, and how it would help people to learn about their energy consumption. Interview II lasted from 10 to 20 minutes, with an average duration of 14 minutes (SD = 5).

4.9 Results

The following section presents the findings sorted by the four parts of our field study. Table 1 lists the 13 participants (P1-P13) from the nine participating households (H1-H9).

H1	H2	H3	H4	H5	H6	H7	H8	H9
P1	P3	P6	P7	P8	P9	P10	P11	P12
P2	P4							P13
	P5							

Table 1. Households and participants in Study 3.

4.9.1 A Priori energy literacy

This section reports the findings from the responses that were given by participants before looking at the recorded data in the first part of the first interview. They gave a variety of answers to the first two open-ended interview questions about which household appliances consume the most electricity and how much. Responses varied from the washing machine (mentioned four times), the fridge (mentioned three times), the shower and the oven (mentioned two times, respectively), the tumble dryer, leaving the lights running or devices plugged in, an electric fireplace, the TV, the computer, and kettle (mentioned by only one participant each). Participants reported low confidence in their responses. Only P12 and P13 (H9) were confident that their electric shower consumes the most electricity. More than half of the sample was unable to quantify electricity usage in any way. For example:

> *'I know what is kilowatts and watts, but that doesn't mean anything to me (P1)'.*
>
> *'I guess that it [oven] might cost 2 Pounds an hour? (P8)'.*
>
> *'The only measure that I can say to gage is the amount of time that is on. So for measuring, I measure by… you know, so... the devices that are on the most (P11)'.*

Three participants spontaneously used Watts or kilowatts as a unit of power. P3 listed his computer screen (15-20W), the Internet router (7W), the electric shower (1,500W), and a lamp in the kitchen with two bulbs (60W). P13 guessed 100W for the baseline consumption, 800W for elevated baseline with lights and computer switched on, and 8kW for maximum consumption. P9 remembered that the baseline consumption in his old flat was around 300W and therefore guessed it should be around 500W in his new place because it is bigger.

We used sketching to reflect users' mental ideas of how much electricity they used over a typical day. All participants opted for a solution with the timeline on the horizontal axis. P12 tried to use the image of a clock in the first place, but realising that a 12-hour clock do not work for a 24-hour day, she concluded that is '*probably easier then to use a graph'*.

The sketches varied greatly in sophistication and detail. P3, P8, and P13 drew staircase-shaped graphs with square waveforms; P1 drew triangle waveforms while the remaining nine participants chose smooth line graphs with sine waveform. The labelling of the x-axis ranged from equidistant time steps in numerical scaling to semantic anchors such as ‘Morning, Midday, Afternoon, Evening’, coinciding with participants’ daily routines. For the y-axis, P1, P6, and P11 did not use labels whatsoever. P2 noted down ‘Consumption’, P7 ‘more elec[tricity]’ by the axis. P 3, P4, and P5 chose ‘kWhrs’ for their shared sketch, P12 opted for kW. P8, P9, and P13 added numeric values to their kW-scales. Figure 9 shows three representative examples of the sketches that are further described in the following paragraph.

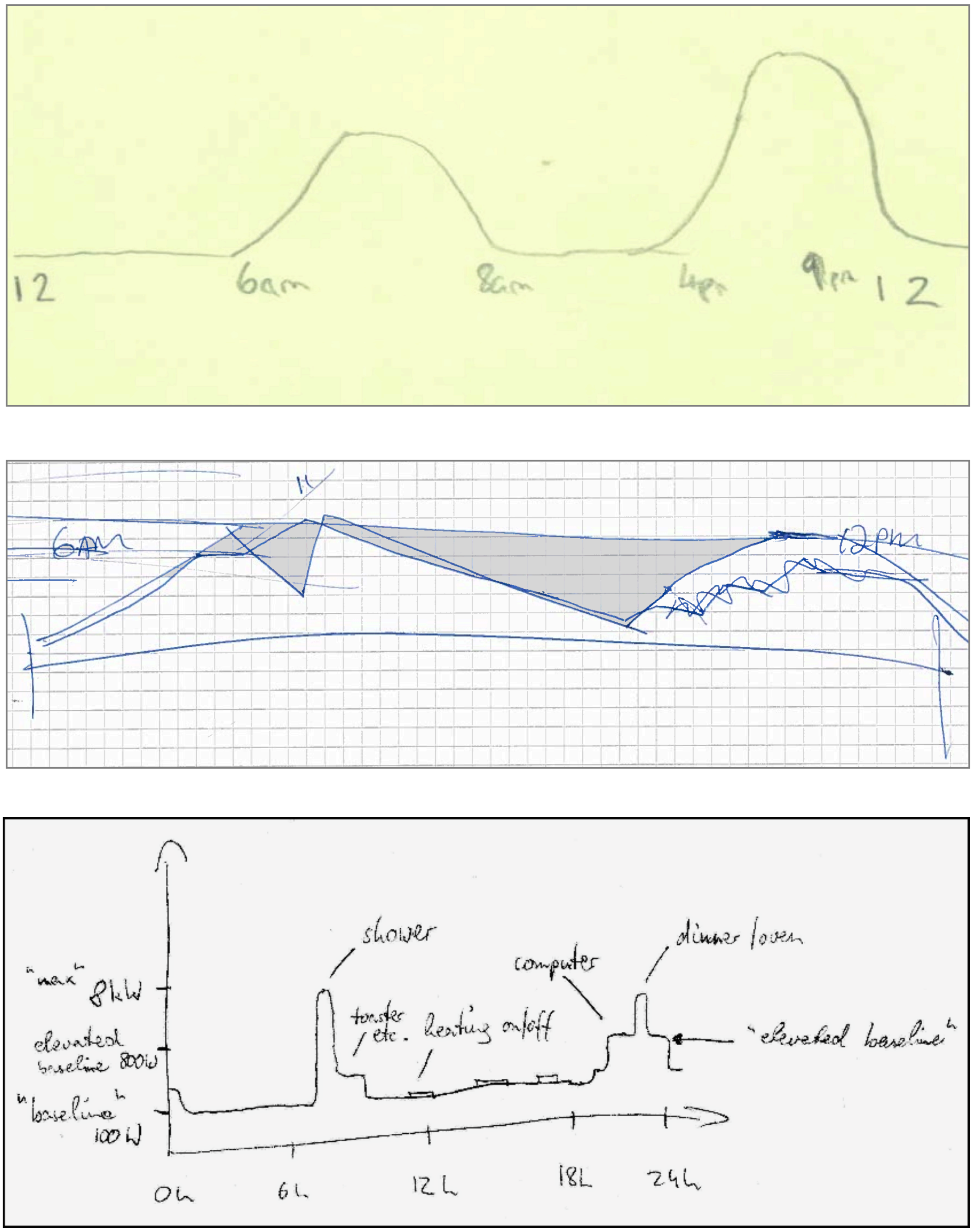


Figure 9. Participants' sketches of daily electricity consumption: a, b, c.

Figure 9a by P11 shows a smooth sine waveform like graph from midnight to midnight. It indicates when the household members are getting up and ready for the day, and when they return home in the afternoon and evening. When being asked about the height of the consumption depicted in her graph, P1, who drew Figure 9b, explained: *'that's not so much a measure of how much, but more kind of going up when we are here'*. Figure 9c by P13 is most sophisticated, with square waveform and labels on both x and y-axis. P13 depicted clean-cut on- and offsets of device usage in his staircase-shaped graph, he quantified how much electricity is consumed by the distinct activities, and he annotated the graph to explain the peaks in the data pattern.

Participants' comments while sketching revealed that what they drew was in accordance with their daily routines. The leading motive in participants' narrative is the time of day and the practice that is typically performed at that time. The curve would rise when getting up in the morning, those with electric water heating might mention taking showers, others making breakfast, including boiling the kettle, toasting bread, making coffee, etc. The curves would drop where all household members had to go to work or children to school. The curve would then rise again in the afternoon when children came home from school, adults returning from work, taking up evening routines such as cooking, charging the phone, and watching TV. P6 assumed that *'at the weekend it's probably high all the time'* since the family was at home. Daily and regular routines would be mentioned more often than less regularly occurring activities such as washing laundry.

4.9.2 Contextual Inquiry

In the second part of the first interview, we would ask the participants to log in to their Loop account and to think out loud (i.e., to verbalise what information they see in the graph and to explain which appliances or activities have led to the displayed patterns). In the following, the results for the emerging themes are presented.

4.9.2.1.1 Routine-Based and Memory Based Reasoning

Our first key finding is that participants had a harder time to account for peaks that were caused by less habitually performed actions - such as washing laundry or vacuum cleaning. Similar to the approach when drawing their sketches in the first part of the interview, participants would focus on explanations that related to typical routines that they perform every day:

> *'It kind of goes up and down throughout the day depending on when we are up and about. It goes up a little bit in the morning when we are... eh... we are all kind of here and doing a few things and the electricity comes on and then it goes down again. And then... it dips when we all leave (P1)'.*

Daily or weekly habits served as a basis for generating ideas about what might have caused the data pattern:

> *'Friday lowest period. Everyone out for a drink in the evening? (P1)'.*

Our second key finding is that participants relied on their memory. They would draw from their knowledge of what they had done recently to interpret the displayed pattern. Therefore, as their memory of events or activities faded with time, the more historical the data the less confident the participants were about their interpretations. Two participants checked their calendars during the interview to help make sense of the data pattern. P8 could not explain a peak and with the backup of his digital calendar, he then described how he had people visiting that day; he was then able to identify the point of their arrival and when he would be cooking roast chicken for the guests. Likewise, P7 checked her pocket calendar to find out whether she was working from home or at the office on a particular day.

4.9.2.2 Interpretation Errors

Interpretation errors are the third key finding. Resulting from the routine-based reasoning, people would commit errors such as assigning peaks around lunchtime to cooking: P8 inferred that his cooking caused a 2kW peak around lunchtime, until he realised a bit later that he had a washing machine running at the same time. P6 tried to understand a big peak in one day when she was at work. First, she considered:

> *'maybe my husband was at home yesterday. I don't know (…) He would have his computer plugged in and be using it. And might also charge his phone cause he doesn't have a routine to charge it. He would be making himself tea and coffee and… stuff'.*

However, the peak was higher and longer than computer and phone would account for, so the interviewer asked what other appliances could have been running in the period of question, which prompted P6 to realise:

> *'Ooh, so that's what that could be as well. Oh, because Tuesday is the day that the cleaner comes. And she comes between 9 and 12. So that's who was at home and she is doing the washing'.*

P3, P4, and P5 (who live together and were interviewed together), spend some time discussion a reoccurring high peak in the morning hours while going through the data of different days, until they finally figured out that the electric shower was causing the peak:

> *'P4: What is that peak?!*
>
> *P5: That is the washing machine.*
>
> *P4: That's 8 o'clock this morning. That is my hair dryer (…) Oh my god, I am not gonna be allowed to dry my hair again (P4 and P5 laugh) (…)*
>
> *P3: That's 7, there, that peak.*
>
> *P5: What were you doing?*
>
> *P4: Drying my hair.*
>
> *P5: That was your hairdryer?*
>
> *P4: Wasn't for very long, though (laughing) (…)*
>
> *P5: Did you wash your hair? Where you late or something?*
>
> *P4: No, I do it every other day.*
>
> *P5: Your hairdryer doesn't show up.*
>
> *P4: No, I didn't use it, I don't do it every day. (…)*
>
> *P3: I got up and had a shower. Maybe my showers are quite hot.*
>
> *P5: Maybe your showers are quite long.*
>
> *P4: Your shower is higher than my hairdryer.*
>
> *P5: It's amazing it uses that much. Just for a shower!*

P4: But… I use the shower before I dry my hair… so it's the shower that's using an awful lot!'

4.9.2.3 Unaccountable Patterns

Several times, participants tried to recall what they did on a given day when explaining the peaks, but concluded: *'I cannot remember'*, staying unclear about what had caused the data pattern. Generally, the longer the day of question dated back, the less confident participants were about their interpretations of the graph's ups and downs. Other than that, there were periods in which the interviewee had not been home and therefore could not have memories (as in the earlier example when P6 puzzled over what her husband could have been doing).

P3, P4, and P5 were surprised that their consumption was higher on Thursday and Friday than on Saturday and Sunday of the same week, although there was one person more in the house during the weekend. Similarly, P9 struggled to comprehend how the usage in the afternoon could be marginally lower than in the middle of the night. P6 could not explain to herself why there would be small increases in usage during the night and why the graph would start rising between 6am and 7am, when none of the household members gets up before 7am. P13 did not come to a conclusive explanation of what was going on during one of his Saturday evenings. First, he reasonably argued that his computer's graphics card must be a significant contribution to the evening consumption because he had been playing a computer game. Then, however, he remembered that they had cooked that night and, thinking more about it, decided that the usage of the evening in question was just the elevated baseline and the computer wouldn't be that high after all.

A design factor of the visualisation that concealed distinct events in the data was the low temporal resolution of the graph (for example the use of the electric shower and the hair dryer in the morning, which would blend into one peak in the graph and prevent the users from detecting the two events). P9 articulated criticism, thinking the tool is *'mostly a gimmick'* leaving them *'frustrated. Not frustrated, too strong a word. But I would like more granular control'*.

4.9.2.4 Disaggregation

Some participants were able to identify big events such as the electric shower or the washing machine and the tumble dryer. Overall however, they performed rather poorly at mapping the data patterns to their everyday actions. They lacked insights about which actions were contributing towards high or low usage days:

> *'It says what was your lowest day and then... what did you do differently that day? I suppose I did go to work (laughing) (P7)'.*

P10 could not make sense of his high evening peak until the interviewer pointed to the fact that the graph displayed global usage and that one peak could consist of several appliances running at the same time. He then listed several devices that would most likely contribute to his big peak around bedtime (he switches on his dish washer and his washer dryer before going to sleep).

P13 concluded that the lack of mapping between data and everyday actions made it difficult to proceed to the stage of changing behaviour:

> *'If I wanted to ehm, actively consume less energy - yes I think it would be helpful to know how much I am actually using. Ehm, it's a little bit like... obviously I know if I turn off the light, I know I use less energy. But it's a little bit like putting things to scale, how much difference does one light bulb compared to the fridge make? That sort of thing would be interesting to know and I think without that sort of software it might be difficult'.*

He added he would be able to work out this information over time. By contrast, other participants repeatedly asked for breakdowns of the information on the appliance level and graphs of activity: P2 tried to figure out which appliances were contributing to a big peak and was looking for a breakdown on the device level on the website, in vain. P3 stated that the software disappointed him, because he thinks it should do the work for the user to *'look at patterns and classify usage of different things'*. He reported having used other personal informatics systems before, so he would have expected graphs of activity. Thereupon P5 added to the discussion with P3 that she could have taken notes:

> *'You know, when I had the washing machine on and the cooker. Which would have been helpful. Cause these are… big things. I would have made a note, had I known it might be more useful'.*

Likewise, P12 said that it is *'a bit like a food diary. It would be a lot of work. Because you have to recall what you were doing'*.

The rest of the sample, when asked if and what they learned from the Loop and if they have ideas how they could change their behaviour to save electricity, would not give a concrete answer. They proposed universal ideas that were not related to what they had seen in their graphs before. They said they would need to look further into it. One person suggested to look at the live-spend widget while another household member is having a shower to determine how much the shower consumes or to look at the graph and read the activity level at times of the day when a certain routine such as making cups of tea is performed.

4.9.3 Long-Term Usage

In the follow-up interviews after three months, there were three participants (P1, P2, and P7) who had never looked at the data again after the first interview. P1 found that the software did not give her anything tangible. P2 did not see any benefit and P7 said she had been interested but had completely forgotten about it. They would have wished for better explanations, for information that was easy to grasp and for a breakdown of the global usage.

Four participants (P3, P6, P9, P12) had looked at their data again in the first weeks after the initial interview, but had quit by the time of the second interview, because they did not gain any further understanding over time. P3 said: *'We were like 'oh okay I am here for the weekend' (laughing). So. It wasn't gaining any insight into electricity usage as such'.*

P6 had looked again when being prompted by her children who wanted to know how they were doing. She described her insights as a *'one-off thing'*. She explains how in the first interview, she had first been puzzled about her data: *'I thought why on earth is Tuesday the biggest day?'* She then thought about it and realised her cleaner was in on Tuesday and she concluded that *'You get that understanding the first time you look at your data and think about it. And I don't think it becomes any more nuanced over time.'*

P9, who had only logged in again to '*kill time*' further stated that *'I find that the information it offers is too general to really give me any inspiration to log back in and continue using it.'* The software does not offer the services he is really looking for, namely information on *'which particular appliances or what particular events might have used more energy'*.

P12 said that by using the comparison function of the Loop (that compares the household to other households), she had learned that *'we actually have a lower resting energy consumption than I thought. We thought our fridge is really bad, but it's actually not that bad.'* She also noted that in a *'more complex household, where you had a lot of appliances, it would be difficult to gage what's doing what.'*

Finally, P8, P11, and P13 were still monitoring their electricity usage with the Loop, where P13 read the weekly digest emails and did not log in to the website anymore. P8 and P11 stated they had been looking every two to three days at times. The consensus was weekly, especially for reading the email digest, with a tendency to check more often when using the app. P13 had unsuccessfully tried to scrape the website for the full dataset.

With these participants, we inquired if they had gained understanding of their data patterns over time and if they had learned more about their household's electricity usage. P8 had learned to identify when his son is home from school, but was not quite sure what is happening at home:

> *'I need to look slightly in more detail but yeah if someone makes himself a cup of tea then you can see that. And he might get home anywhere between 4 o'clock and 6 o'clock. So... but as soon as he gets home he turns on the computer. But I am not sure, the computer isn't that visible'.*

When asked if the data had helped him learn about his household, P11 said:

'A little bit. I think it could be more micro. I'd like to get down to the nuts and bolts. I'd like to know about (…) zonal areas where it actually shows you, kind of in a visualisation of your house, showing you heat maps of where the usage is taking place so you can quickly zone in on these areas, as opposed to me having to turn things on and go back turn things off (…) I wouldn't say I've become better. I could pre-empt when (…) electric usage was gonna happen (…) I think with it being the peaks of the entire house... if it could be more detailed. I think that'd be better'.

One specific insight he had gained over time was related to the electricity usage of his teenage daughter:

'Yeah, the peaks and troughs (…) in the diagram. Looking at what has been used over the particular day and when (…) I can do as I say turn to the rooms and be "hang on a minute there must be something left on somewhere". And I can actually trace those things (…) her room is slightly... you know, the epicenter of all electricity usage'.

P13 commented on his monitoring over time saying he *'was interested a little bit in how it would develop. So with more data, because in the beginning fluctuations were high, and then after several months you can say, okay, so this is our weekly average.'* Further, he had gained understanding about the relative consumption of the washing machine and the electric shower over time:

'So about the first interview where I was surprised that the washing machine draws more electricity that the shower. Now, that is because of the integration period. So the actual amount when the shower is on is much higher, but then we don't shower for 90 minutes. While the washing machine runs for 90 minutes. Because of the binning of the graphs it looks like the washing machine draws... I mean the total amount of energy is higher. But the peak amount is less. So yeah. Given that some thought'.

He had also consulted the Loop data when he re-negotiated his direct debit with his energy provider. Although his bills and the Loop data were not completely congruent, he used the Loop to get a better idea what he was spending on electricity every month. He answered the question if he had learned from the data feedback by saying

‘I put some thinking into it which otherwise I wouldn't have. And it gave you some insights (…) Insights like the washing machine, the shower.’

4.9.4 User Requirements

In the second part of the follow-up interview, we asked participants which characteristics and functions would be important to them if they could design the ‘perfect smart meter feedback’. The functions that were requested most often were first, appliance-wise data disaggregation to make the information more actionable, and second, interaction with the software to explore the data.

In reference to the Loop’s line graph, P3 pictured the following visualisation:

> *‘you could have like an aggregate of how much power you're using. But then underneath that, you can have other lines or bars or some sort of visualisation showing 'ok so this is what was contributing to that much (…) you could maybe have a list of appliances. And you could roll over that appliance and then it would go from being like a greyed-out line to being a high-contrast line. So you can see what that particular appliance did over the week or over the month. Or over the day’.*

P9 brought up a similar idea:

> *‘say having baseline usage, devices that run all the time, colored in one color, and when something new starts, that that takes on a new color. And so when another device starts that becomes another color again. So you end up with like a stacked graph’.*

Others had the idea of a screen that would show a schematic flat or house or the actual property with its rooms. The display would then show the consumption per room, and per room they would be able to zoom in and obtain more details on the device level, such as their efficiency and how they could be improved:

> *‘I'd like to know about (…) zonal areas where it actually shows you, kind of in a visualisation of your house, showing you heat maps of where (…) the usage is taking place so you can quickly zone in on these areas (P11)’.*

One imagined a little display by every light switch and in the display every appliance would be represented with a little picture (of the appliance) that contains information about it.

One suggestion included a smart home system that would integrate information from the Internet and offer tailored advice; every plugged-in device would automatically communicate its specifications as well as its system status to the network, and thereupon the user could be sent useful notifications. Similarly, another participant suggested that whenever there is a new appliance, there could be a training phase for the system to learn to recognise all appliances. As an example of specific tailored advice, P9 suggested the system could *'tell you that this particular light bulb is using more than the other ones. Or that your TV uses 25% more energy than most people's TVs; did you know that you could save energy by… upgrading this device or using it in a different way or something'*.

4.10 Discussion

4.10.1 Main Findings

We asked participants which appliances or activities in their homes consume most electricity and how much electricity these consume. We observed a three-way split between participants with very accurate knowledge including quantitative specifications, participants with reasonable guesses, and participants whose guesses were inaccurate. The sketches granted an interesting insight into participants' energy literacy and mental models of how they thought electricity would be used over the day. As expected from the literature, there is a relationship between people's sketches and their understanding of data (Walny et al., 2015). The details of the graph and the ability to label the scales revealed that our sample ranged from very low to very high literacy.

Compared to Study 2, the sketches allowed a richer insight into participants' energy literacy, and the sketching exercise gave participants more time and degrees of freedom to think about and choose their scale labelling (rather than being put on the spot by the question which unit they would use to measure power or energy). The question asking participants to quantify their consumption could only be answered by participants with very high energy literacy, and not at all by those with low literacy. The open-ended question which appliances use most electricity is very limited in that the response can only be classified in a binary way as reasonable or not. For example, the response 'leaving the lights on' reveals one specific misconception, but it doesn't reveal the participants' knowledge about all other activities and appliances in the home. A ranking task including multiple appliances generates more data and allows for more nuanced scores.

Three months later, we observed the same three-way split between those that had quit using the tool immediately after the first interview, a group that had tried but failed, and one group that was still using the tool. That some of our participants stopped engaging with the Loop soon after installing it is consistent with prior work that has also shown that some users abandon smart technologies within the first weeks or months (Harrison et al., 2015; Lazar et al., 2015).

In terms of understanding why a person might choose to continue to use a device, Hekler, Klasnja, Froehlich, and Buman (2013) discuss moderating variables that influence how efficient interventions are for different people. Moderation is important in behavioural theory because research needs to address key differences and cater for different user needs or motivation. We found that only sufficiently energy literate users would continue using the Loop, while more illiterate participants could not be motivated to keep up the tracking. For the literate ones, the drivers seemed to be curiosity and fascination with the data (Epstein et al., 2015; Rooksby, Rost, Morrison, & Chalmers, 2014). Like Hargreaves, Nye, and Burgess, (2010), we observed interesting dynamics between parents and children in the household, where it was sometimes the children that nudged parents to engage with the feedback out of curiosity.

4.10.2 Reflection

The results of the contextual inquiry give insights into how people interact and reflect on time series energy graphs. We learn that all participants sketch time series graphs – hence it's an intuitive visualisation. However, we found that participants did not find this type of data visualisation useful when exploring their recorded usage data. The ability to make sense of the data feedback depends greatly on the individual's analytical skills (Kempton & Layne, 1994) and participants often struggled to understand and explain peaks and troughs in the graphs. The Integration stage (where information is processed for the user to reflect on it) in the Model of Personal Informatics (Li et al., 2010) is automated and done by the software. However, the Integration must serve the Reflection phase, meaning that the collected data needs to be processed and visualised in a way that facilitates and catalyses reflection and gain of knowledge. The data should be visualised in a manner that is clear and easy to analyse (bottom-up).

Instead, we found that participants relied on top-down processes (i.e., they relied on their memory of what they were doing at specific times to help explain patterns of use). The implication of this is that people are often biased by 'active events' and so possibly overlooked energy-intensive but less routinely performed activities – e.g., washing laundry – or background devices that consume electricity not specifically tied to an event – e.g., the fridge. These findings are in line with previous work where participants would mostly look at peaks in the graph, neglecting the baseline (Costanza et al., 2012) or overlooking practices that they thought were non-negotiable (Strengers, 2011b).

Concerning the waveform, the staircase-shaped square form (as in Figure 9c) best represents the real-world matter (Pinker, 1990), as it shows clean-cut on- and offsets (mind that P13, the most literate participant in our sample, chose the square wave format). The smooth sine waveform (as in Figure 9a) reflected lower literacy in the pre-interview, and in the contextual inquiry the sine shape turned out to be a problem as it is impossible to see beginning and ending of an event, and worse, distinct events blend over into one another.

The most prominent theme in the user requirements interview was disaggregation. Everyone demanded a view that would break down usage per appliance, or at least render separate streams visible in the time series display. Appliance-wise disaggregation would allow users to reflect on the usage information more easily, thus eliminating the problem of using top-down processes in the interpretation. By showing aggregated energy data current generation smart meters are preventing people from transitioning to the stage of meaningful Reflection and Action. Froehlich, Kay, Larsen, and Thomaz' (2014) description of personal informatics failures related, among others, to problems regarding the data collection and regarding the user interfaces. In our case, the approach for collecting and displaying aggregated data with low frequency does compromise comprehensibility of the graph. We assume that this very mapping is crucial because users reason in terms of everyday actions and educational approaches should take relevant routines and situations into account (Álvarez & Vega, 2009; Hargreaves, 2011). The action- or event-based nature of thinking about energy consumption is not mirrored in the data, so people fail to map data patterns to behaviour and to gain relevant insights. It is assumed that disaggregated feedback would be more actionable for householders, but research has not yet delivered strong evidence to support this assumption (Kelly & Knottenbelt, 2016).

We aimed to sample more than one participant per household (Yang et al., 2015) and we interviewed all three participants from Household 2 together which revealed an interesting dynamic. Discussing the data together brings the advantage of combined knowledge and the discussion may stimulate the sense making process. On the other hand, the disadvantage is that it is harder to determine the individual sense making capacities. Interviewing participants together might increase ecological validity. The follow-up revealed that many households differed with regards to who looked at the eco-feedback and how much the matter was talked about between different household members.

4.10.3 Limitations

This study is exploratory field work with the purpose of examining the problem space and generating more specific research questions regarding the cognitive sense-making processes in interpreting electricity data. We aimed for a mixed sample to see how different types of users would read and reflect on the data. Yet, there are several limitations that we address in the following.

The sample is relatively small, but it is consistent with that seen in previously published qualitative research (e.g. Yang, Newman, & Forlizzi, 2014). The contribution is an in-depth understanding of how people understand energy data. A qualitative research approach is appropriate to address this research question. As opposed to quantitative data where certain sample sizes are required for the validity of statistical tests, an increase in sample size is only useful as it reveals additional themes in the qualitative data. In their guide for qualitative research, Blandford, Furniss, and Makri (2016) argue that a pragmatic approach should be taken to recruiting participants. Participants should be recruited until 'theoretical saturation' is achieved (i.e., the point at which gathering and analysing more data on the chosen theme does not yield further insight). Saturation was achieved in our study. For example, only one out of 13 participants came up with an alternative idea for sketching her daily consumption (and in the end opted for the timeline as everyone else did). The analysis of the interview data revealed reoccurring topics between participants (e.g. memory based reasoning and disaggregation as the most prominent ones). Given the considerable overlap between participants' data, we are confident that we have reached saturation in observing the cognitive processes involved in householders making sense of the Loop data feedback.

We could not interview all participants in person. All are UK households, but in different cities, and some of the participants had very busy schedules and found it inconvenient to meet in person. To respect participants' wishes and due to limited resources for travelling far, we chose to interview some participants via Skype given the practical constraints. The data between Skype and personal interviews is comparable and we have no reason to assume that they had different experiences from one another.

Finally, we have investigated how people reason about the usage data displayed by the Loop and possibly learn from it, albeit we did not record actual behavioural measures. It was beyond the purpose of the study to explore if people would reduce their consumption. Although we can confirm that users ask for appliance-wise feedback, it remains to be investigated if they'd perform better with disaggregated data.

4.11 Conclusion

The study suggests that cognitive information processing must be given more consideration in designing energy feedback. Simple line graphs seem suitable to visualise energy data as power over time, but overall, users fail to link this visualisation to practices in everyday life. Individuals vary in their ability to interpret the graph and independent of energy literacy and graph literacy, the line graph visualisation is very prone to errors caused by heuristics and false memories. The low temporal resolution and the inability of the tool to record the impact of practices in the home limited the usefulness of the feedback. Despite the rich data history, little insight is provided as to: Which practices consume most energy? How could consumption be reduced most easily?

4.12 Introduction to Study 4

Study 3 found that a simple energy data visualisation, a line graph showing power over time, was not well suited for householders to reflect on their consumption and their practices. A major problem was that events (i.e. appliances) could not be identified in the graph. Participants relied on memory or heuristics to explain peaks, which is error-prone. One participant in Study 3 suggested she could have kept a diary to help her reflect on the data on display, and across participants, appliance-centric information and interactivity were named as user requirements. Study 4 reports a study with an interactive feedback tool that allows householders to annotate the history of their energy consumption. Instead of engaging with the feedback in a one-off interview, participants in this study were asked to use the tool for three weeks and to keep a digital diary by annotating the data, labelling peaks in the graph with the practices they've been attending to at the time.

Interaction is considered an important element in Info Vis, triggering reflection in users and allowing them to become more engaged and to actively explore the data (Munzner, 2014; Prost et al., 2015). Schwartz et al. (2015) have introduced the idea that it is key what people do with technology, as opposed to what technology does to people. How people engage with their data is crucial because it determines whether they think for themselves and that enables them to decide if they want to improve their behaviour (Ploderer et al., 2014). Despite the many benefits of AI and smart solutions, there is a fine line between automating smart systems and sustaining users' active engagement (Alan et al., 2016). Interactivity seems a good solution for smart systems to keep the user in the loop.

Costanza, Ramchurn, and Jennings (2012) have developed an interactive software prototype called FigureEnergy. FigureEnergy allows householders to annotate and manipulate a graphical representation of their residential electricity consumption to reflect on their usage pattern and to learn how, when and to what end they use electricity (Figure 7). That means, they can select time periods in the graph and label them 'breakfast', 'washing machine', etc. The idea is that users link the data to their everyday social practices (Hargreaves, 2011) through the interactive annotation. Users can annotate the graph while they are carrying out a household task and see how the graph rises in near real-time (reflection-in-action) and they can look at their usage profile at the end of the day or at the end of the week (reflection-on-action) (Schön, 1987). FigureEnergy aims mostly at reflection-on-action, which allows for more extensive interaction and experimentation with data than merely observing changes in real-time. Active manipulation of and reflection on data have been found to be effective for behaviour change (Fogg, 2002) which can be explained by constructionist learning theory (Papert, 1980), which assumes that learning is a process in which the learner actively constructs their knowledge.

The FigureEnergy publication from 2012 (Costanza, Ramchurn, & Jennings) describes the details of FigureEnergy's original version's development, implementation, and technical evaluation. Study 4 presents a new study that uses FigureEnergy as a probe to understand in more detail how users reflect on their consumption with the help of an interactive feedback tool that provides a rich data history.

Study 4 describes the deployment of a revised version of FigureEnergy in the field with twelve participants from nine households who were asked to use it for three weeks and to annotate their electricity consumption in FigureEnergy. The annotations that participants made in FigureEnergy over the three weeks were analysed. At the end of the three weeks, participants took part in an interview which explored how they had used FigureEnergy, how they reflected on their data, if they learned about their energy consumption and if they changed any of their practices. As in Study 3, multiple participants took part in the interview and were interviewed together (Yang et al., 2015). All interviews were audio recorded, transcribed and the transcripts were imported into the qualitative data analysis software Nvivo and analysed thematically (Aronson, 1995; Clarke & Braun, 2014; Seidman, 2013).

4.13 Method

As outlined under 'Collaborations' in the preface to this thesis, Study 4 was designed and conducted by Enrico Costanza (collaborator at UCL) who kindly shared the collected data with the author of this thesis, who conducted the data analysis. The Method section nonetheless describes the Material and Procedure for the clarity of this document.

4.13.1 Sample

The advertisement asked if the reader wanted to learn more about their domestic energy consumption with the help of an interactive software prototype. Nine households were recruited and provided with electricity consumption sensors and the FigureEnergy software prototype (further described in Material). Twelve participants from nine households (one or two participants from each household) took part in the interviews (eight females), all from a suburban area of London. The average age of participants was M = 54.4 years (SD = 14.3 years). The households ranged from flats to terraced houses with between one and four bedrooms with between one and five occupants living in the home. Seven households owned their home.

4.13.2 Material

The sample was provided with the Web application prototype FigureEnergy. The data sensing relied on off-the-shelf digital networked electricity meters. FigureEnergy has two main views, the Consumption Graph and the Consumption Overview. The Consumption Graph (Figure 10) displays the recorded electricity data as time series line graph showing average power use in the home. The task for participants was to annotate this line graph, thus inviting reflection. Users can navigate through the graph in time and zoom in and out. They can select a time period using the mouse, e.g. 7pm-8pm, and then annotate this time period by adding an ‘event label’, e.g. ‘meal dinner’. FigureEnergy comes with a set of event labels such as ‘meal breakfast’, ‘toaster’, ‘kettle’, ‘computer’, ‘washing machine’ and so on. Participants can use these labels by selecting the provided icons for these labels. Further, FigureEnergy allows the user to remove, manipulate, or complete these suggestions by adding textual descriptions to describe in more detail what participants were doing and which appliances they were using.

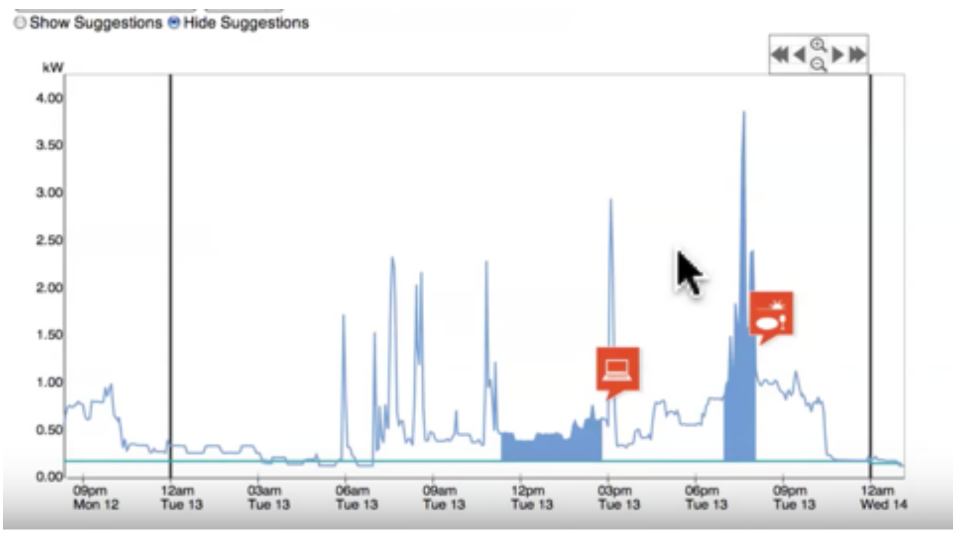


Figure 10. FigureEnergy's Consumption Graph.

The other feature of FigureEnergy is the Consumption Overview (Figure 11). Based on the annotation in the Consumption Graph, FigureEnergy calculates the energy per event label and displays it as an energy-centric, or event-centric, two-dimensional rectangular box with an area proportional to the energy consumed by the event. The purpose of this view is to shift away from the time-centric display of the Consumption Graph and to emphasise the amount of energy consumed for a certain appliance or activity, thus allowing for deeper reflection-on-action by analysing patterns in the data. By doing this, it should be easier for users to relate to the intangible energy information and to compare events and to see easily where they are consuming a lot of energy. For example, does watching TV consume as much energy as washing laundry over the course of a week? By hovering over the boxes, users could retrieve additional information, such as the textual descriptions they had added, kWh consumed, and duration of the event.

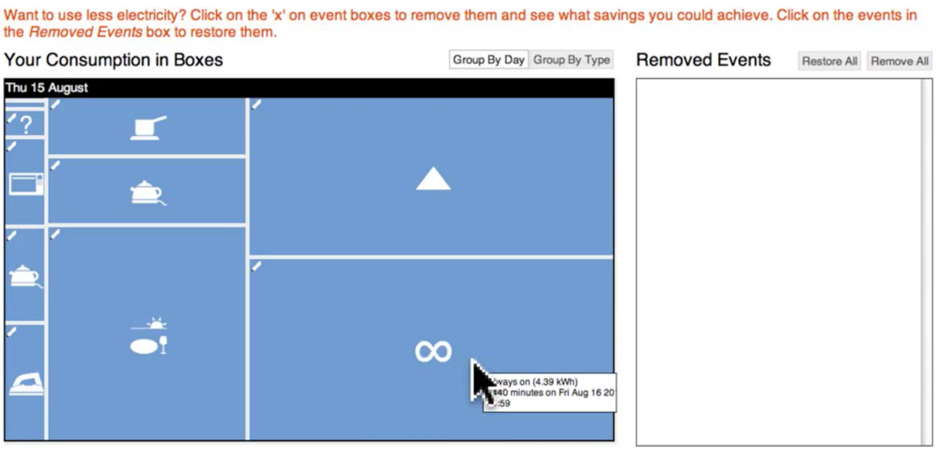


Figure 11. FigureEnergy's Consumption Overview.

4.13.3 Procedure

The study started with an initial home visit where the electricity consumption sensor was installed and the use of FigureEnergy demonstrated to each participating household. Participants were asked to access the system daily to annotate the peaks in their consumption data. After approximately three weeks (depending on availability), a follow-up visit was arranged to conduct a semi-structured interview and to collect the sensor kit. The interviews were audio recorded and fully transcribed. They lasted from 50 to 108 minutes, with an average duration of 67 minutes (SD=19).

4.14 Results

First, we present results on how often participants accessed FigureEnergy and how many annotations they made during the three-week deployment. The annotations were assessed for plausibility. Then the thematic analysis of the interviews is presented. Table 2 lists the twelve participants from the nine participating households.

H1	H2	H3	H4	H5	H6	H7	H8	H9
P1	P2	P3	P5	P6	P7	P8	P10	P12
	P4				P9	P11		

Table 2. Households and participants in Study 4.

4.14.1 Annotations

4.14.1.1 Access Frequency and Number of Annotations

Participants' access to FigureEnergy is reported in Figure 12. Access logs from the server were processed to identify interaction sessions (a session was defined as an access period where the user was not inactive for more than 5 minutes). The orange line shows the total access to the Consumption Graph across all participants. The grey line shows the total access to the Consumption Overview. Throughout the study, the Consumption Graph page was accessed considerably more than the Consumption Overview page. Moreover, the frequency of access to the system went down over time.

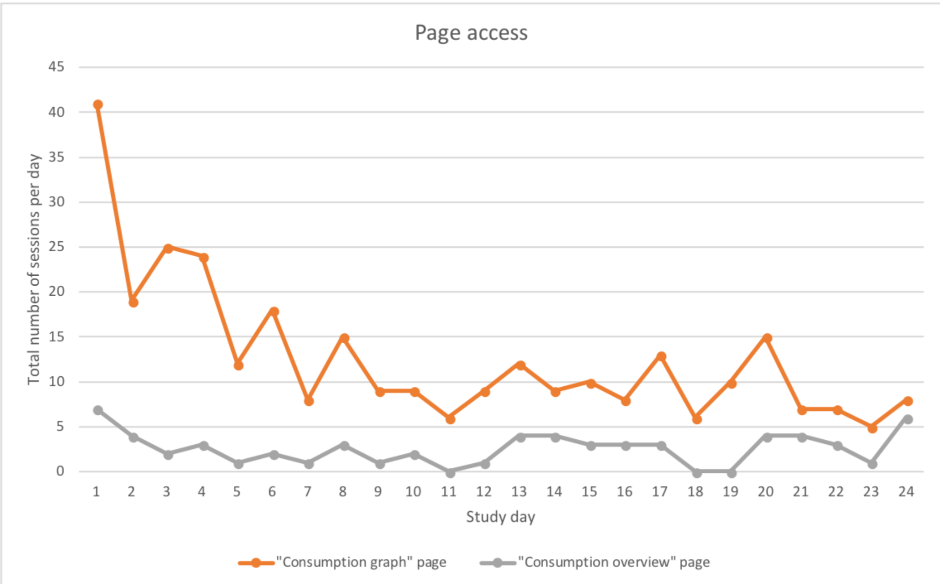


Figure 12. Participants' page access.

Participants created a total of 1,054 annotations over the three weeks of the study, corresponding to an average of 117 annotations per household (SD=77.0), and 5.6 per household per day. Figure 13 illustrates the number of annotations generated by each participant per type. The data is reported in two groups: annotation types related to specific appliances (such as computer, dishwasher, kettle) are listed on the top of the figure, while annotations that are more generic (such as housework, meals, or 'other') are listed on the bottom.

	H1	H2	H3	H4	H5	H6	H7	H8	H9	Total annotations per type
Air cooling	2									2
Computer	1	1	11	16	1	1	6		4	41
Dishwasher	4		3		3	6		2	3	21
Heating	5			10	1	1	10		4	31
Hob					2				2	4
Ironing			8		3		1		5	17
Kettle			2	54	24		49	14	23	166
Lighting	5		1	29		11	44			90
Microwave			4	3	1		11	1	2	22
Oven	7			1		5		7	15	35
Showering and hair-drying	7		2	14	15	13	23		1	75
Toaster					1	1	1	8		11
Washing and drying	2		11	4	8	21	17	18	7	88
Watching tv	4		7	46	16	12	83	25	14	207
DIY							1			1
Gardening								2		2
Housework	2		1	1	6	2	3			15
Meal breakfast		2		14	8	10		18	2	54
Meal dinner		2		13	6	18		14	2	55
Meal lunch		10		6		9		1		26
Other A		18	1	7		1	16		15	58
Other B		12		2		1			4	19
Other C		11	1	1	1					14

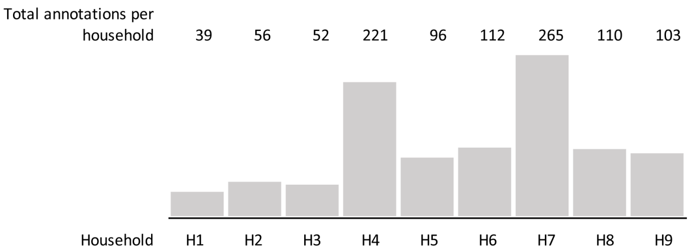
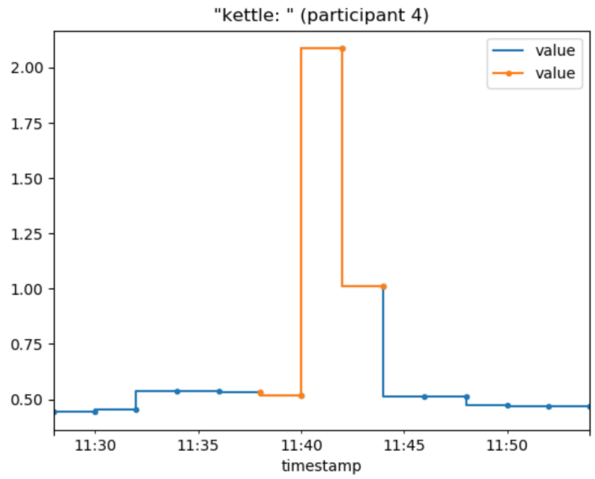


Figure 13. Number of annotations per type and per participant.

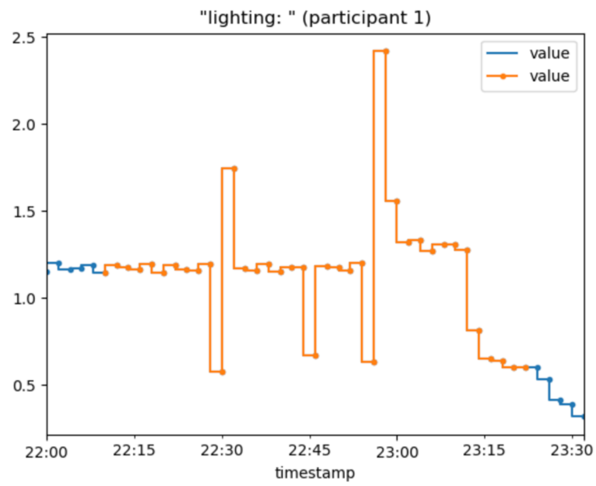
4.14.1.2 Plausibility of Annotations

The annotations were coded independently and iteratively by the author of this thesis and a second researcher (Enrico Costanza, collaborator at UCL) until consensus was reached. Each annotation was coded as 'correct' or 'incorrect' according to whether it was considered plausible or not, based on duration, energy amount and power pattern. Figure 14 shows an example of an annotation that was coded as correct (a peak of 2kW that was labelled 'kettle'). Figure 15 shows an example of an annotation that was coded as incorrect (an interval of oscillating power up to 2.5kW labelled as 'lighting'). Figure 16 shows the percentage of correct annotations per type and per participant. General annotations (e.g. housework) were counted as correct, hence the higher accuracy of the generic labels on the bottom.



Type:	kettle	Description:		Start:	2 Oct 11:38
Duration:	8 minutes	Mean Power:	1.04kW	Total Energy:	0.14kWh

Figure 14. An example of a 'correct' annotation.



Type:	lighting	Description:		Start:	15 Oct 22:10
Duration:	74 minutes	Mean Power:	1.14kW	Total Energy:	1.41kWh

Figure 15. An example of an 'incorrect' annotation.

	H1	H2	H3	H4	H5	H6	H7	H8	H9	Overall accuracy per annotation type
Air cooling	0%									0%
Computer	0%	100%	73%	100%	0%	100%	100%		75%	85%
Dishwasher	75%		100%		0%	100%		100%	100%	81%
Heating	20%			5%	100%	100%	40%		38%	29%
Hob					25%				100%	63%
Ironing			75%		67%		100%		80%	76%
Kettle			0%	85%	58%		98%	93%	74%	83%
Lighting	0%		100%	97%		91%	41%			63%
Microwave			75%	33%	0%		91%	100%	50%	73%
Oven	86%			100%		90%		100%	87%	90%
Showering and hair-drying	43%		0%	57%	63%	85%	100%		100%	74%
Toaster					100%	100%	100%	100%		100%
Washing and drying	50%		64%	75%	63%	100%	97%	94%	71%	86%
Watching tv	50%		100%	91%	19%	83%	80%	96%	64%	79%
DIY							0%			0%
Gardening								100%		100%
Housework	100%		100%	100%	83%	100%	100%			93%
Meal breakfast		100%		100%	100%	100%		100%	100%	100%
Meal dinner		100%		100%	100%	100%		100%	100%	100%
Meal lunch		100%		83%		100%		100%		96%
Other A		78%	100%	86%		100%	100%		87%	88%
Other B		92%		100%		100%			100%	95%
Other C		100%	100%	100%	0%					93%

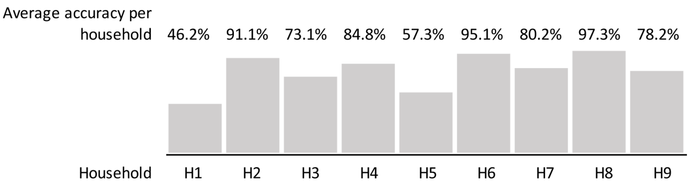


Figure 16. Correct annotations per type and per participant in percentage.

4.14.2 Thematic Analysis

In the interview, participants reported how they had been annotating their Consumption Graph. We first present the themes that emerged around annotation styles. Then themes are presented with regards to the reflection and learning that took place based on the interaction with FigureEnergy.

4.14.2.1 Peaks versus Baseline Consumption

P10 and P11 (H8) said that what they labelled most were the kettle, the toaster, the washing machine, the tumble dryer and the TV. Indeed, these are their most frequently used event labels (Figure 13). P6 (H5) said she focused on the peaks in the morning when getting up and among her most frequent annotation types were the 'kettle' and 'showering and hair-drying'. Sometimes in the interviews, participants referred to events that they did not annotate because they did not translate into peaks. P1 (H1) for example spoke about the radio in the interview: she explained she leaves it on all the time for her dog (and her partner questions if that is necessary). When asked if she ever annotated the radio in FigureEnergy, she negated, explaining that *'It doesn't surge'*. She referred to it as constant usage and said she would not know how much appliances that are on in the background like the lights or her computer are using but she said it would be interesting to see a breakdown. P5 (H4) thought it would be good to single out the fridge which is always on in the background.

4.14.2.2 Practices versus Appliances

Around food practices, FigureEnergy provided the event labels 'meal breakfast', 'meal lunch', and 'meal dinner'. P11 and P10 (H8) reported that they would sometimes annotate specific appliances used to prepare food (i.e. oven 7 times, kettle 14 times, toaster 8 times, microwave 1 time; Figure 13), other times they would use the 'meal' event labels (i.e. breakfast 18 times, dinner 14 times, lunch 1 time). P11 (H8) admitted that he used the meal labels when he was *'pushed for time'* because it was easier than specifying which appliances he used. P5 (H4) elaborated further on whether he would use activities or appliances for his annotations: for example, if he made a cup of tea for breakfast first and porridge in the microwave later, he would annotate the two appliances separately. If he did everything at once he would choose the breakfast symbol.

P5 (H4) mentioned it would be nice to be able to break everything down into the separate appliances that might be used for making the meal. He explained that a label like 'making porridge' would be ambiguous because he could be using the microwave or the hob to make it. P1 (H1) never used the 'meal' event labels suggested by FigureEnergy and explained that they *'don't really eat like that'*: she might only boil the kettle to make a cup of tea in the morning and when she cooked in the evening she would use the label 'oven'; she points out that *'you don't use any more electricity while you're having your meal'*. P6 (H5), like P5 (H4), said it was *'tricky'* to annotate when multiple things were on at the same time and she would have liked a button in the software to annotate multiple items. She used the breakfast label to summarise *'the toaster, the kettle and maybe the oven'* and similarly she used the 'washer dryer' label indiscriminately for washing and drying. P7 (H6) described a period when she was cooking dinner and running the washing machine and the dish washer and her partner had a bath, so she *'just sort of put one thing [label/icon] and I listed the others* *in it [the textual description]'* and she *'couldn't really segregate what was happening in terms of usage with which, because it all seemed to be happening at the same time.'* P12's (H9) annotation style was distinctively different from the others': she consistently selected time periods over many hours and wrote a list of all things she did and used in that time.

4.14.2.3 Lighting

An appliance that stood out were the lights. Over all participants, lighting was the second least accurately annotated type of event. At the same time, it was the third most annotated event. What did participants say about lighting in the interviews? P6 (H5) mentioned that she never annotated lights and now wonders how they and other appliances in standby (her multiple fridges, the TV, the alarm and chargers) add up. P10 (H8) and P11 (H8) neither used the 'lighting' label for their annotations, nor did they speak about conventional lights in the interview – except that they mentioned that the light in their aquarium has high wattage. P7 (H6) said she can see *'quite clearly when people get up'* because there is a small increase in electricity consumption when the lights are switched on so P7 (H6) did annotate the lights. P12 (H9) said in the interview that she noticed a *'slow rise'* when she switched lights on; she did list the lights twice in her long annotation lists. P8 (H7) and P9 (H7), too, annotated the lights when they noticed a *'little blip every morning'*. Yet, their annotations for light only reached 41% accuracy (Figure 16).

4.14.2.4 Surprises and Mysteries

Five of the participants reported having *'surprises'* or making *'discoveries'*. P11 and P10 (H8) as well as P8 (H7) described FigureEnergy as an *'eye-opener'*. P11 and P10 (H8) remembered that the biggest peaks were caused by the dishwasher, the washing machine and the tumble dryer. P11 (H8) also reported being surprised how high the kettle and toaster spike. P1 (H1) and P8 (H7) found out that the electric shower consumes more electricity than they thought. P8 (H7) and P3 (H3) learned about the power needed for ironing. P8 (H7) said that the iron causes a peak in electricity usage which had never *'crossed [her] mind'* before. P3 (H3) said she always believed that *'anything that heats really drains your electricity'* and found this assumption confirmed for the iron. Yet, P3 (H3) *'could not believe (…) the enormous spike'* her hairdryer caused. P6 (H5) said at one point she was surprised by the height of her baseload, which is caused by her having three fridges and freezers. P8 (H7) stated that *'when you think about electricity you don't think about the fridge freezer'* but using FigureEnergy made her think that she will consider energy efficiency ratings when the time comes to replace the fridge freezer, realising that this is *'going to be a big percentage of your bill'*. P12 (H9) also reported how she looked at the data together with her husband and it reminded them how the house is still *'alive'* even when it's *'empty'* because the fridge freezer and other appliances are on standby still using electricity.

Occasionally, participants referred to *'mystery'* events in their data, both in their annotations as well as in their interviews. For example, P4 and P3 (H3) talked about spikes that they could not explain and therefore had not annotated. In the interview, they referred to them as *'unknown'*, *'in bed'* or *'out'*. P2 (H2), too, reports *'a couple of unexplained spikes'* that she labelled *'no one in'* or *'don't know'*. P5 (H4) noticed *'tiny little things'* during the night and wondered if it can be the fridge freezer. P1 (H1) mentioned *'random patterns (…) at sort of three o'clock in the morning'* and she had no clue what that could be. P6 (H5) was *'surprised'* and *'puzzled'* that even when nobody is around there were still fluctuations and peaks.

4.14.2.5 Reflection on Self-Reported Waste

A theme that emerged in participants' reflections was to do with self-reported waste, such as keeping appliances like the TV on standby out of *'laziness'* (P10 and P11, H8). P2 (H2) and P1 (H1) talked about keeping appliances on for their pets. P1 (H1) runs the heating for herself and her dog, whereas her partner gets too warm. Hence, they sometimes run the heating and the fan at the same time. Interestingly, the events annotated as *'air cooling'* by P1 were coded incorrect. Yet, she critically reflected on this practice. When the dog was unwell, P1 (H1) ran the heating during the night and her partner opened the window. P1 (H1) also uses the heating to dry her clothes and sometimes keeps it on all night so that her partner's work clothes are dry in the morning. She commented on these anecdotes in a way that indicates that she considers them wasteful *('That's bad use of electricity!', 'That's really bad isn't it?', 'excessive'*).

P2 (H2) has annotated in FigureEnergy that she leaves a hall light on during the night, and in the interview, she explained:

> *'I'll get up during the night and go to the loo and that way I can see where I'm going basically and I don't have to put the big light on in the bedroom which is more likely to then wake me up a bit more (…) And also (…) my cat is not allowed in my bedroom at night because she keeps me awake, so being the big softie I am I leave the hall light on for her, just completely ridiculous I know'.*

Similarly, she mentioned that she has started leaving the radio on for the cat. She reasons:

> *'I mean people would say, do you need to leave a light on for a cat [laughs], no, clearly I don't, but it's my choice to do that. So I…I know that I'm using energy for that, but that's my choice and I'm paying for it'.*

In this case, the wasteful behaviour is explicitly related to personal choice, justified by the payment of the energy.

P3 (H3) spoke about another case of keeping lights on, which she has not annotated in FigureEnergy. In her case, it is the outside light in front of the house – *'I know it seems really wasteful having it on all night'* but otherwise she would struggle to find her keys and it's also a security question for her because it makes it look like someone is home. In contrast, she referred to herself as the *'electricity police'* switching everything off when not needed, because *'waste generally irritates'* her and she has always *'hated waste'*. Her heuristic is *'if you don't need it, you shouldn't really be using it'*, explaining this is *'an environmental thing'* and *'general awareness'*.

P6 (H5), when thinking about wasteful behaviours, explained that they have a TV in each of the bedrooms and she and her partner and her daughter sometimes all watch different programs simultaneously while her son plays the Xbox. She reasons they could vote on a program and all watch together, but other than that, she do not see how they could save as they are out during the day (this reflection only came up in the interview, the annotation data did not include any evidence of this).

Despite these reflections about waste, participants reported that they are *'reasonable energy efficient'* (P2, H2) and *'only use what's necessary'* (P1, H1). P7 (H6) and P3 (H3) said they already only wash clothes if they are dirty and that they only wash full loads of laundry. Relating to cooking practices, P7 (H6) *'multitasks'* when using the oven trying to use the heat to cook several things.

When asked about how they could save energy, P1 (H1) suggested, theoretically speaking, using less lighting, not listening to the radio all day, and wearing jumpers to have the heating on less. P7 (H6) considered to maybe precook meals on the weekend and reheat them on weekdays, and jokingly said they *'could eat more salad'*. P6 (H5) reckoned they *'probably do waste power (…) like any family'* but she did not think she would do anything differently because she did not consider them being very wasteful *('I don't think we're that wasteful')*. P6's baseline was high due to the three fridge freezers in the house. When asked about getting rid of one or two of these appliances, P6 (H5) explained *'you hear people, their freezers break down and so at least with having the three, the three, you know, if there was a problem with one we could then swap it into another'*. She said if there were *'financial restraints'* they *'could cut that down maybe'*. In contrast to resistance to change, the next subsection reports instances where participants mentioned behaviour change.

4.14.2.6 Self-Reported Behaviour Change

P8 (H7), who had described the feedback as *'quite an eye opener'*, stated that she has become concerned about her consumption. She and P9 (H7) reported a range of insights and consecutive behaviour change, for example P8 (H7) has *'been more frugal with the use of the dryer since doing this [taking part in the study]'* and she switched the lights off more often as opposed to having them on during the day. At times, they had been running an electric heater in their daughter's room but upon discovering how much it consumes they reconsidered using it and concluded warmer pyjamas would do. Equally, they used to put their daughter's towel in the dryer *'just quickly to warm it up when I got her out the bath'*, which *'made [them] think that it was a pure (…) luxury rather than a necessity'*. They pointed out that the information did not make them say *'that's got to stop'* but rather got them to think. P9 (H7) reasoned:

> *'Well, you got to have lights on when it's dark, you got to have your fridge on. You've got to make a cup of tea now and then, you know, you've got to have a shower. There's things that you just can't avoid, but there are a lot of things you can avoid (…) Oh, or the other thing is, you know, if we do boil a kettle, where in the past we might have boiled it, walked away and then ten minutes later re-boiled it. We don't do that anymore, you know. You kind of think, actually if we're going to make a cup of tea, make a cup of tea and then go and do whatever is going to distract you'.*

P12 (H9) found that the study was *'making you aware again because I think you do get complacent'*. She had occasionally turned the radio off completely instead of keeping it on standby. She had made further changes regarding the washing of dishes and clothes:

> *'I thought no actually today I'm not going to put [the dishwasher], I'm going to wash the breakfast things up, I'm going to wash the lunch and wash the evening meal things up (…) it has been a conscious effort thinking no I don't think I need to use [the dishwasher] as much as I do (…) I think I could live without the dishwasher (…) I think we got a little bit lazy with it. You know, we just fill it up and we use it on a daily basis. I think I would reduce it down to probably just the weekends'.*

For washing her clothes, P12 (H9) reported washing less now than she used to and using a more efficient program:

> *'I have put the washing machine on quick washes as opposed to longer cycles in the last two and a half weeks (…) I said to my youngest daughter, "You've only worn these jeans today." She's worn them all day at work fair enough, so I think really you could wear them sort of two or three times as opposed to putting them in the washing machine because they're not dirty, because you've only sat in them all day in the office. You know, you could air them (…) they're not dirty. And they smell fresh still. You can still smell the comfort on them for goodness sake. I'm sure you can wear them tomorrow'.*

P4 (H3) admitted that in the past he would sometimes *'be ironing, having the telly on, have the laptop on, stop ironing for a bit, answer a couple of emails or something* *like that but you've left the iron going at the same time'*. P4 (H3) had learned to make changes based on what P3 (H3) learned from the feedback:

> *'One of the things P3 said to me, and I'm conscious of now is, "don't turn the iron on and iron one shirt. If you're going to turn the iron on, you know, iron quite a lot of stuff because otherwise you're going to get a great big spike"'.*

In contrast to participants' accounts about changing practices, the annotation data did not provide apparent evidence that they really changed the practices.

4.14.2.7 Personal and Generational Circumstances

The study seemed to have sometimes triggered reflection that goes well beyond the energy consumption data that participants were presented. Participants referred to a variety of factors, such as their upbringing, generational issues, financial matters, socio-economic comparisons, trust or mistrust towards utilities, convenience, comfort and self-reported waste.

P5 (H4) for example seemed to use energy quite reasonably (for example, he washes his clothes only after wearing them several times and his showers are as quick as three minutes), he was not concerned about cost, and yet he was very mindful of getting the best deal, shopping around for the best tariffs *('I suspect that when that contract expires, they would...they would try and probably, you know, push me up. And that's the point when I think, again, almost like in principle, no, you know, you aren't going to bully me. I'll shop around'*). Reasoning about whether he would change his behaviour to save energy, P5 (H4) did not think the technology was going to change his lifestyle:

> *'If I want a hot drink I'll have one (…) I'll put the kettle on. If I want to, well, use some electricity in some way, I will do it when I want to do it (…) I'm 80 years old for Christ sake, I haven't got another 40 years to go. Whereas if I was in my 40s, and the children were still at home, I would be far more concerned possibly and have an opportunity to do something about it. Now I can afford to be if necessary a little bit reckless (…) I've been fairly lucky in life (…) just down the road in reality there are people who are probably not even as old as I am who are retired, and will be reluctant to put on heating because*

they're worried about the cost. I mean I think that is a factor in a lot of people's lives'.

P12 (H9) thinks that her daughters are from a generation that thinks *'everything is just automatic'* and *'They don't really question'* it. She explained how she grew up in the 70s with *'shortage'* and *'power strikes'* whereas her daughters just plug things in without realising what that means in terms of energy consumption. P12 (H9) said *'for my eldest daughter, she's not a silly girl by any means but I suppose she just thought jumping in the shower was using water – not electricity.'* Equally, this daughter *'doesn't like the house quiet. So the telly's on even if she's not in the room'*. Similarly, P8 (H7) said that as a teenager she would put a pair of jeans that she wanted to wear in the washing machine and that she *'would never think of doing that now'.*

P7 (H6) stated that she is very conscious of electricity usage, feels the responsibility towards her children's generation and she *'worr[ies] about the planet'*, elaborating that *'it's all about being responsible to…being responsible for our planet and all the creatures that live on it, not just ourselves, and just being a good person, really.'* She feels guilty about the increased consumption compared to her parents' generation and finds that everyone needs to be more responsible: *'You use more power for more things these days than, than we had when we were children, so it is only increasing and yet the resource is only decreasing.'* She described this as a culture in which we are using a lot and that it should be a comprise between wanting things and using too much. P7 (H6) tried to get her kids to switch things off when not needed: *'I feel I am permanently saying…coming down in the morning and saying, "Why is the house lit up like a Christmas tree?" and if you've got something charging turn it off when you've finished charging it'.*

4.15 Discussion

4.15.1 Main Findings

The results of this study show that participants did rather well in terms of explaining their energy consumption data patterns. This finding is contrast to the observations from Study 3, in which many participants could not explain their energy consumption using feedback provided by the Loop system. One reason that FigureEnergy might be superior to the Loop is that it allows users to annotate events that contributed to spikes in energy consumption. The annotations that participants made over the course of three weeks were judged to be mostly very plausible with exception of two participants whose annotation accuracy fell under 60%. Interestingly, the participants with the lowest annotation accuracy also mention some of the most energy-inefficient behaviours in the interview (such as heating and cooling the house at the same time and having three fridge freezers) yet they do not necessarily recognise them as inefficient (the three fridges were considered necessary). In contrast, participants whose annotation accuracy is above 70% seem more likely to be either more economic or to identify and change wasteful habits.

One might think that annotation accuracy is high for participants who were environmentally aware and economical to start with. The cases of P8 and P9 (H7), P12 (H9), and P3 and P4 (H3) invalidate this assumption. Schwartz et al. (2015) have introduced the idea that it is key what people do with technology, as opposed to what technology does to people. For example, P8 and P9 (H7), whose annotation accuracy is good, used to engage in highly energy-intensive behaviours before taking part in the study, but upon reading in the data and recognising their profligacy, they reported they stopped using the tumble dryer when not necessary (e.g. for preheating pyjamas). Participants who achieved high annotation accuracy were more likely to learn from the technology, to recognise profligate use, and to change accordingly.

The findings are typical in that participants were still more likely to misjudge certain appliances. As usual, lights stand out as an appliance that is annotated frequently, but often incorrectly (Attari et al., 2010). At the same time, it is an encouraging finding that several participants learned that lights cause only a very small increase in the graph. The results are further in line with Study 3 with regards to patterns that participants could not explain and referred to as mysteries. Participants using Loop encountered fluctuations, often during the night but occasionally during the day, that they had no explanation for.

Consistent with the findings from Study 2 and 3, the results of Study 4 confirm that data on the appliance-level is crucial for householders. Occasionally, participants pointed to the difficulties of annotating the graph when multiple events here happening simultaneously in the home. They also mentioned that it was not clear how baseline appliances (like the fridge) were contributing and they would like an automated breakdown. While FigureEnergy was running with a single sensor collecting total consumption, the idea of the prototype was that users would annotate events, and then the software would calculate the energy consumed by each event and display this information in the Consumption Overview. Alas, participants missed the opportunity and did not engage with the Consumption Overview during the three weeks (Figure 12) and did not talk about it in the interview.

The purpose of FigureEnergy was to provide a rich data history to users and to make the energy feedback as activity-centric as possible. Focusing on social practices in the home, event labels suggested by FigureEnergy included, amongst others, the option 'meal'. The focus on a 'meal' (as a practice) as opposed to for example the oven or the hob (as an appliance) was meant to facilitate the annotation and to link the data to social practices. However, the practice-centric labels limited the usefulness of the label because it made annotations ambiguous. For example, the label 'meal' could refer to a range of different appliances. It was further pointed out that only preparing a meal would consume energy (whereas the practice of having the meal would not). This is against the previous assumption in the original FigureEnergy study that activity-centric feedback matters more than appliance-centric information (Costanza et al., 2012). It seems that appliance-specific information might in fact be better suited for the data feedback because it is more specific and leaves less room for interpretation and thus reduces the work load for the user to ascertain which appliances they used.

4.15.2 Reflection

The core difference between Study 3 and 4 is that thanks to FigureEnergy as an interactive prototype, participants could annotate and actively reflect on their data patterns as often as they wanted, and they were encouraged to do so daily. In contrast, participants in Study 3 were asked to reflect on their data as a one-off activity. The advantage of FigureEnergy (Study 4) over Loop (Study 3) can be explained by constructionist learning theory (Papert, 1980) and other works that emphasise the importance of reflection (Schön, 1987). By actively engaging with and reflecting on the energy consumption of their everyday actions, participants were able to construct new knowledge. Most participants annotated events on the same day when they had taken place, which reduced the negative effects of false memories and heuristics (which were apparent in Study 3).

The results provide novel evidence of how people read and reflect on energy data. We found two patterns of 'reading energy data': we refer to them as reading in the data and reading beyond the data. By reading in the data we mean householders who are analytically reflecting on the energy data. A couple of participants referred to FigureEnergy as an eye-opener (so did participants in Hargreaves, Nye, and Burgess (2010) study on energy monitors) and several identified the appliances with the highest consumption in the home. This is valuable because knowing where the energy goes is the first step in reassessing one's energy use. It might be interesting to focus on the householders who learned and to understand how they learned, i.e. to focus on the 'bright spots' of the intervention to learn from the successful cases (Harrison, Bird, Marshall, & Berthouze, 2013). The most valuable information for participants was appliance-centric information that they used to reassess their practices.

Some participants were reading beyond the data: they talked about aspects to do with generational issues, upbringing, financial matters, socio-economic comparisons, environmental concern, mistrust towards utilities, convenience, comfort and self-reported waste. It is important to note that reading beyond the data, as opposed to reading in the data, is not indicative of either good or bad energy use or of poor annotation accuracy. Participants who went beyond the data tend to have high annotation accuracy rates, and they also produced a comparatively high number of annotations, i.e. reading beyond the data does not seem to replace reading in the data, it complements and extends it.

‘Waste’ is an interesting example of going beyond the data because it always goes beyond the objective data. Whether a practice is considered profligate is inherently subjective (Schwartz, Stevens, Ramirez, & Wulf, 2013), because needs like cleanliness, comfort, and convenience are strong motivators (Shove, 2003). How economically energy is used is typically shaped by childhood education, comfort preferences and material circumstances (Strengers, 2011a). Whilst ‘waste’ is highly subjective, this is problematic because the potential to save energy is bigger for householders with high-consumption profiles. To help users reflect about waste, energy feedback could support people by providing nudges and personalised recommendations to make feedback smarter in the future (Mogles et al., 2017).

4.15.3 Limitations

The sample was relatively small and predominantly female which restricts the generalisability of the findings, as gender has been found to impact how people respond to energy feedback (Hargreaves et al., 2010). It’s also likely that novelty effects of using the system would wear off over time. At the same time, some of the benefits of FigureEnergy are valuable one-off insights (e.g. discovering that the tumble dryer consumes a lot of energy) and do not need to be tracked permanently. Nonetheless, more research is needed to learn how to keep users engaged with smart energy feedback long-term.

Like the previous studies in this thesis, the focus of Study 4 lies on investigating how householders interact with and reflect on their energy consumption data. This study presents self-reported behaviour change only, it does not focus on actual consumption data. It was not possible to tell unequivocally from the annotation data whether the reported behaviour change really took place. Some cases of reported behaviour change were certainly due to the Hawthorne effect (Landsberger, 1958), meaning participants behaved differently simply because they were part of a study (for example, some mentioned FigureEnergy increased general awareness and they switched appliances off when they went away, which was not related to specific insights they had gained from the data feedback).

4.16 Conclusion

The findings of this study suggest that energy feedback can be made more meaningful and efficient through interactive systems that engage users and trigger reflection. Participants in Study 4 were overall quite good at annotating their energy usage data. This contrasts with Study 2 and Study 3, in which participants were widely unable to relate the energy data to events that contributed towards consumption. This discrepancy between the studies suggests that FigureEnergy's interactive annotation feature successfully engaged participants, enabled active reflection, and helped them to understand their energy consumption better. Some participants reported that they identified wasteful practices and they subsequently changed their behaviour.

4.17 Conclusion from the Interview Studies

The research questions addressed by Chapter 4 were: How do householders interact with smart electricity feedback? Do they understand it? Can they link the data to their everyday lives?

Study 2 revealed that some participants chose to not interact with the SM IHD at all. At most, we found participants using it as an ambient feedback tool to keep an eye on their consumption not becoming too expensive during the winter. This was enabled by the near real-time feedback. What the IHD lacked was a richer overview over consumption other than the instantaneous snapshot.

Study 3 in comparison used a web-based tool which provided an extensive record of electricity consumption over time. This allowed for deeper interaction and reflection, but participants struggled to identify the causes of spikes in energy use and it did not help them understand how they were using electricity for everyday practices or how they could rethink consumption.

Study 4 addressed the lack of interaction by introducing an interactive annotation feature which succeeded in triggering a reflective analysis of participants' data and even beyond the data. Effective reflection centred around appliance-level information, with users connecting the data patterns to everyday practices and identifying where they were using energy (and where they were using it profligately).

Across the three studies, the findings confirm the central role of reflection in eco-feedback. If the purpose of feedback it to change users' behaviour, there is a process that people must go through which involves a stage of understanding the information. If this stage is neglected, i.e. if the feedback design neglects to design for understanding, users will stop at this stage and not transition any further and they will not know how to change their behaviour even if they want to.

Chapter 5 Visualising Disaggregated Electricity Data

5.1 Introduction to the Visualisation Studies

Chapter 5 addresses RQ 3: What are the effects of disaggregation and data visualisation on householders' learnings? Does disaggregated data help householders to learn more about their energy consumption than aggregate data? How should disaggregated data be visualised?

There is a general assumption that appliance-level information will be more effective than feedback about the total energy consumption of a household. This approach has intuitive appeal and is considered by some to be the solution to giving better home energy feedback (Armel et al., 2013). However, a recent systematic review of the results of several studies that have deployed smart meters that give disaggregated feedback found limited empirical evidence that this approach led to any more savings than simply giving aggregate feedback (Kelly & Knottenbelt, 2016). The lack of evidence in favour of disaggregated feedback is not evidence that appliance-level feedback is not useful to people. In fact, the studies reviewed by Kelly and Knottenbelt all suffered from methodological biases and do not allow for valid conclusions about the value of disaggregated energy feedback.

Many energy feedback studies have been conducted in the field to increase their level of ecological validity. The difficulty is that internal validity is harder to control for. This presents a challenge in identifying which of the many potential factors explain the variance in the data (e.g., the timing, frequency, modality of feedback). Singling out the most effective factors across studies with different designs is challenging. Also, effectiveness is usually quantified as the decrease in energy consumption. In terms of behaviour change theory, there is little experimental evidence of how energy savings are moderated by users' data comprehension because there is little research about users' cognitive analysis of the (often graphically visualised) feedback (Chiang, Natarajan, & Walker, 2012; Ford & Karlin, 2013; McCalley & Midden, 2002).

The findings from Study 4 suggest that appliance-specific information should help users to meaningfully reflect on smart energy feedback. Participants in this study were annotating their consumption graph to help them reflect on how much energy was consumed by different practices in the home. Some identified practices that were very energy intensive, and not necessary, and they decided to discontinue these practices. The appliance-level information helped end profligacy by providing actionable insights.

The purpose of Studies 5 and 6 is to find evidence as to whether disaggregated data is more useful to people than aggregate data and to investigate how best to visualise appliance-level data. Automatic disaggregation remains a technical challenge, but it is likely that sooner or later, disaggregated data will be available (either through Non-Intrusive Load Monitoring (NILM) or future Internet of Things (IoT) solutions where smart appliances can communicate their data to a smart home hub, just like SMs currently communicate data automatically). What is still missing is the proof of concept that all the effort invested into obtaining appliance-level data is not misplaced. Demonstrating behaviour change is beyond the scope of this thesis, but the following studies aim to test whether disaggregated data is superior to aggregate data for helping people to make sense of domestic electricity consumption data and improve their energy literacy.

5.2 Introduction to Study 5

Study 5 consists of three lab experiments which investigate different data visualisations for how they enable participants to learn how much energy domestic household appliances consume. The three lab experiments are presented as Studies 5.1, 5.2, and Study 5.3.

5.3 Introduction to Study 5.1

It has been established that householders often have a poor understanding of how much energy appliances consume, and that they systematically misjudge certain appliances (Attari et al., 2010; Chisik, 2011; Kempton & Montgomery, 1982). Energy feedback needs to correct these misconceptions for householders to make better decisions about energy use. Understanding electricity consumption requires understanding the concept of power consumed over time, which is a difficult cognitive task for most people (Kidd & Williams, 2008). Furthermore, energy feedback often uses graphical representations. This adds another challenge, because people are not trained to understand complex data and charts (Baur et al., 2012; Cleveland & McGill, 1984; Tufte, 1983). A critical question that is investigated here is how best to visualise electricity consumption data on the appliance-level to help people learn and retain an understanding of how much energy appliances are consuming.

Energy data is most commonly visualised as time series line graphs (Costanza et al., 2012). Time series line graphs, like the visualisation in the Loop (Figure 8) or the Consumption Graph in FigureEnergy (Figure 10) show power over time. A simple line graph can only represent one measure and omits information due to aggregation (Loorak, Perin, Kamal, Hill, & Carpendale, 2016). Kelly and Knottenbelt (2015) collected appliance-level data and Figure 17 shows a graphical representation of the time series data for several appliances in a household. This visualisation was created in the context of NILM research, not in the context of energy feedback. However, it shows multiple colour-coded line graphs, one per appliance, just like participants in Study 3 suggested when they were asked how the Loop feedback could be improved and made more useful.

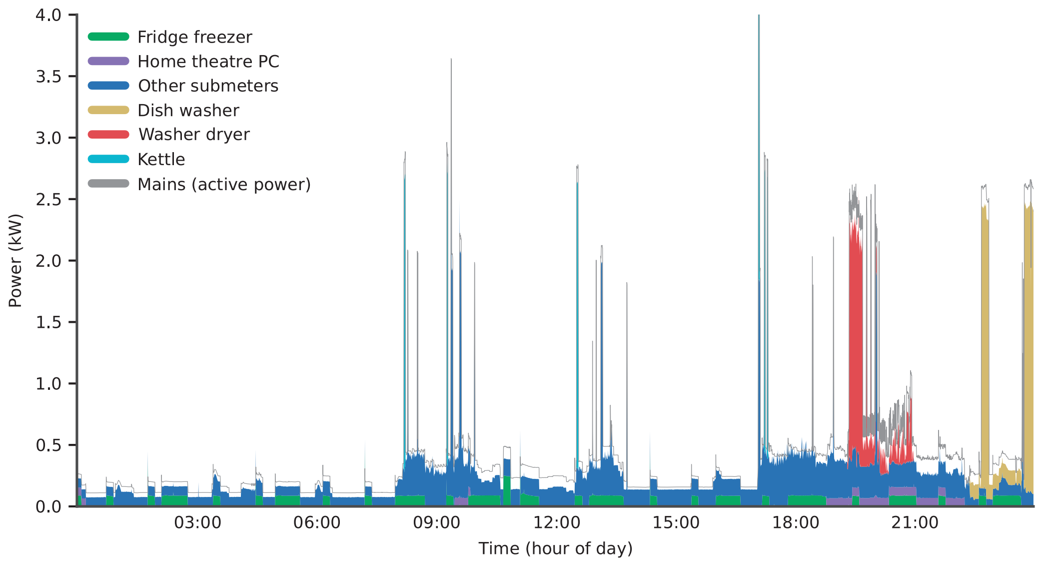


Figure 17. Time series visualisation by Kelly & Knottenbelt (2015).

The assumption that disaggregated feedback is better suited than aggregate feedback to convey information about appliances has intuitive appeal. However, one challenge is the same for both time series graphs: to understand how much energy was consumed, one must estimate the area under the curve. Due to the varied shapes of appliances power consumption patterns, this is a difficult visual task. People are generally very good at detecting deviations from the horizontal (i.e., to process the obvious peaks), but not at integrating power over time (i.e., the area under a line graph) (Tufte, 1983). For example, a kettle will run for a short period of time, using a lot of power per unit time. Whereas a dishwasher will run for much longer, using less power per unit time. Using a line graph visualisation that shows energy usage as a function of time, it is difficult for users to determine the cumulative energy usage of a given appliance over time, potentially making it difficult to determine which appliance uses more cumulative energy over a standard usage cycle.

Commercial solutions that offer a breakdown on the appliance-level typically offer summaries of energy consumption per appliance. Energy monitor provider Voltaware provides visual feedback in the form of a pie chart (Figure 18). Bidgely, the provider used in Sokoloski's (2015) study comparing IHDs to disaggregated and web-based feedback, lists the energy share of appliances in percentage and visualises it as a bar chart (Figure 19).

Figure 18. Pie chart by Voltaware.

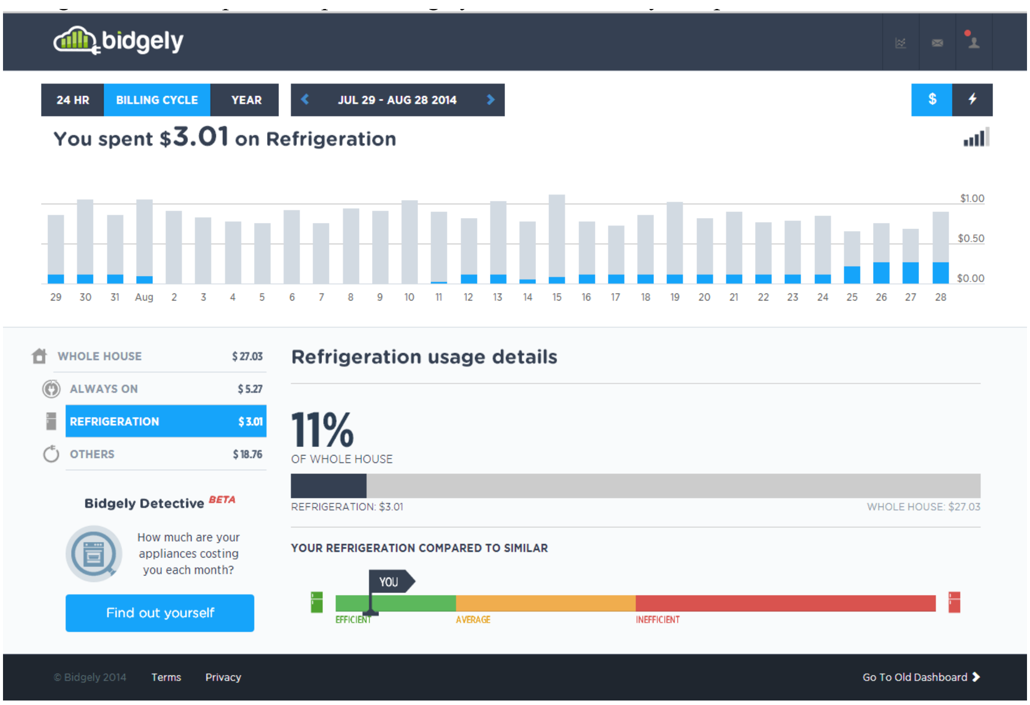


Figure 19. Bidgely web-page from Sokoloski (2015).

The purpose of Study 5.1 is to test whether participants' knowledge of how much energy everyday practices consume changes after being exposed to different data visualisations. In a between-subjects design, the experiment tests three visualisations: (1) an aggregated time series line graph, (2) a disaggregated time series line graph and (3) a normalised disaggregated visualisation that deemphasises time (Figure 22).

To measure knowledge about how much energy everyday practices consume, participants took an energy test, which was based on the ENLITEN energy game (http://www.cs.bath.ac.uk/enliten/) (Lovett, Gabe-Thomas, Natarajan, O'Neill, & Padget, 2013). In this game, participants had to indicate which one of two practices consumes more energy (e.g. making coffee versus running the dishwasher, Figure 23). This game is a useful measure because it tests in a playful way whether people know how much energy is consumed by everyday household practices. After being exposed to different visualisations, the accuracy score on the game (the number of correct comparisons) indicates whether the visualisations lead to different performances in the different groups.

The hypotheses of the experiment are that participants in the disaggregated group will perform better on the energy game than participants in the aggregated group, and that participants in the normalised group will also perform better on the game than participants exposed to both aggregated and disaggregated time series graphs. Study 5.1 is designed to test the assumption that the normalised visualisation facilitates the comparison of how much energy practices consume by eliminating the time factor. On the other hand, the disadvantage of the normalisation is that it takes away all time related information and the specific power patterns. It was the spikes in the Consumption Graph in FigureEnergy (Figure 10) that seemed to have triggered participants' reflection *beyond* the data in Study 4 (4.15.2). Therefore, the assumption is that the aggregated and disaggregated condition might trigger more reflection that the normalised condition.

To assess whether participants reflected beyond the data in Study 5.1, they were briefly interviewed at the end of the study and asked to describe how they made sense of the visualisation they saw. This approach follows Peebles, Ramduny-Ellis, Ellis, and Bonner (2013) approach of measuring understanding. Peebles et al. would ask participants to describe 'something interesting' they saw in a graph, and ask participants a set of semantic comprehension questions that were tailored to the content of the graphs. A similar approach was taken in this study to get a better understanding of how participants interpreted the different types of data visualisations. All interview data was audio recorded and transcribed in the transcription software f5. The transcripts were coded and analysed in Word MS Office.

5.4 Method

5.4.1 Sample

Forty-three participants (12 male) were recruited through the UCL Psychology Subject Pool. Ten participants were aged between 18 and 20 years, 31 were between 21 and 35 years, and two were 36 years or older. All were adults with normal or corrected to normal vision who were accustomed to reading from left to right and who pay their utility bills (or do so with the help of their partners or fellow tenants). Participants received course credit or a small payment for taking part in the study.

5.4.2 Materials

The experiment was designed to see whether participants' assessment of electricity consumption of common household appliances is affected by the design of energy data visualisation that they use. Three energy data visualisations were used: a line graph with a single aggregated data line (representing total energy usage across multiple appliances), a line graph with multiple disaggregated data lines (representing energy usage for each of the individual appliances), and a disaggregated graph that has been normalised over time (representing the total energy usage of an appliance over a single usage of that appliance).

Both line graphs (Figure 20 and Figure 21) show time series data. Duration of usage is represented on the x-axis as time in minutes and electricity consumption is represented on the y-axis as power in Watts. Figure 20 is a line graph that shows how the aggregated power consumption of three different appliances (a kettle, a vacuum cleaner and a dishwasher) varies over time.

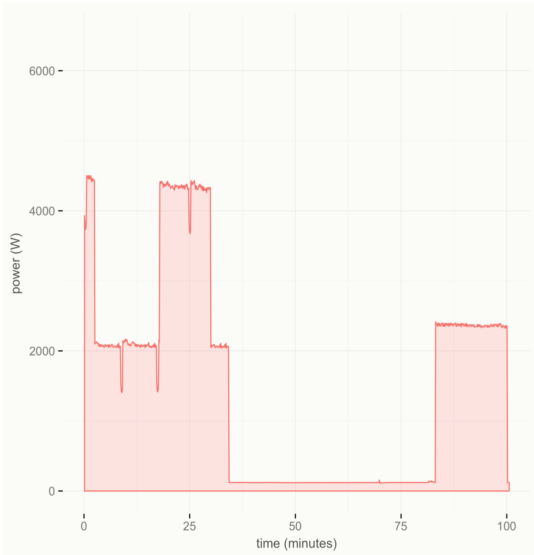


Figure 20. Study 5.1, aggregated visualisation.

In contrast, Figure 21 shows the same data but here the power consumption of these three appliances are represented as different coloured data lines. The intention of the disaggregated line graph is to make it easier for the user to distinguish how the power consumption of each appliance varies over time throughout a period of usage. For example, the dishwasher has a distinct pattern with two peaks and a period of lower usage in the middle. This pattern becomes visible in the disaggregated condition but is invisible in the aggregated condition. The disaggregation will make certain comparisons relatively easy (e.g. the kettle consumes obviously less than the dishwasher), but the more alike two appliances are, the harder it is to estimate the area under the curve correctly (e.g. dishwasher and washing machine).

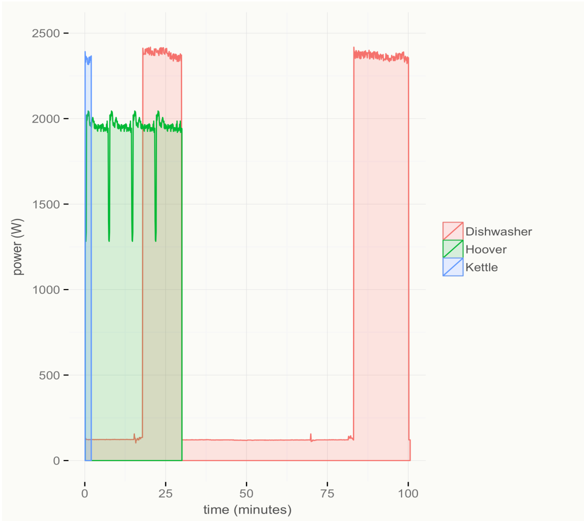


Figure 21. Study 5.1, disaggregated visualisation.

Figure 22 shows the normalised visualisation. This is essentially still a line graph showing energy consumed as the area under the curve, only that time of use has been normalised over all appliances. The intention of this visualisation is to eliminate the challenge of comparing the area under the curve of differently shaped power consumption patterns. The normalised visualisation shows cumulative consumption over a single usage of the appliance. This allows the user to readily see which of the appliances is using more energy over a standard usage cycle.

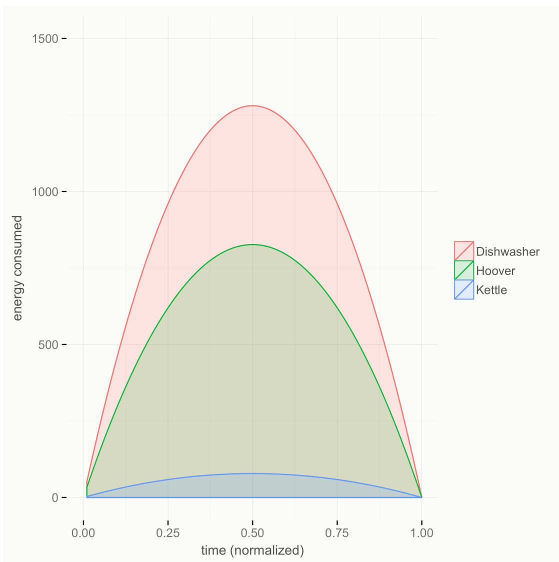


Figure 22. Study 5.1, normalised visualisation.

To assess participants' judgment of electricity consumption we used an energy game, which was a two-alternative forced-choice task (Figure 23). Participants had to indicate which of two practices consumes more electricity during a standard usage cycle, for example, making coffee or running the dishwasher. Every two-alternative forced-choice comparison was shown to participants with the according pictographs of the appliances used (e.g. coffee maker and dishwasher) and the information what it was used for (making coffee and running the dishwasher) and for how long (15 minutes and 1 hour 30 minutes) (Figure 23). For each pairwise comparison, we recorded response accuracy and response time in seconds. In addition, we also assessed response confidence by asking participants how confident they were about their decision on a scale from one to five (one being low confidence, five being high confidence). The pairwise comparison task and the icons that participants click to indicate their answer were based on the ENLITEN energy game (http://www.cs.bath.ac.uk/enliten/).

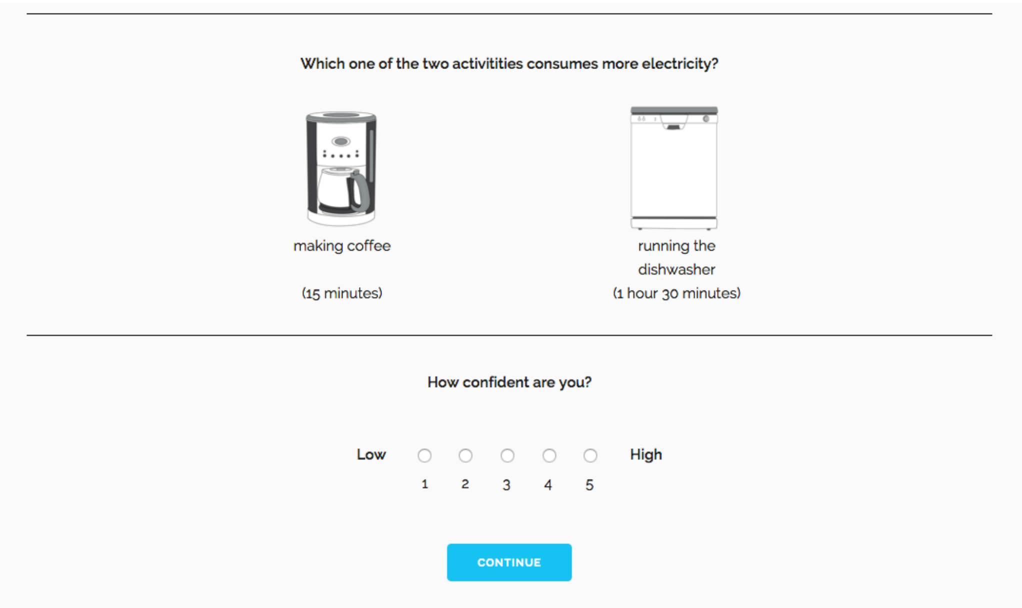


Figure 23. The energy game.

For making both the energy visualisations and the pairwise comparison in the energy game, we used the same set of nine common household appliances. These appliances were: radio, lamp, microwave, toaster, kettle, coffeemaker, vacuum cleaner, washing machine, and dishwasher. To model the energy consumption of these appliances we used data from the UK-DALE dataset (UK domestic appliance-level electricity, Kelly & Knottenbelt, 2015). This open-access dataset is from a study that recorded domestic appliance-level electricity at a sample rate of 1/6 Hz in five UK houses, with the longest recording lasting 655 days in one house (House 1 in the dataset). We use data from House 1 (a London end-of-terrace Victorian house, built around 1905). For each appliance, we identified the typical duration of use and the power usage over time. All materials were presented on a 27-inch iMac (2560 x 1440, Graphics: ATI Radeon HD 4850 512 MB).

5.4.3 Design

The experiment is a pre-test post-test between-subjects design, in which the independent variable was the graphic representation of the electricity data feedback. The dependent measure is participants' knowledge about the electricity that is being used for different practices in the household (such as making coffee or running the dishwasher). We measure the change in knowledge for the nine appliances from pre- to post-test in the energy game. This is measured by response accuracy, response confidence (on a scale from one to five), and response time (in seconds) in the energy game. In addition, qualitative data was collected at the end of the experiment when participants were asked how they had made sense of the visualisation they saw.

5.4.4 Procedure

Participants were informed that they would be taking part in a study about domestic energy usage and they were randomly assigned to one of the three visualisation conditions. Participants completed the study in a small private office with a desktop computer on a desk. The office was quiet and free from external interruptions and distractions.

Throughout the experiment, we used 'Jack' as a persona to embed the experiment into a story about residential smart metering. Participants were told that Jack and his family live in a London end-of-terrace house and that they have a Smart Meter in their house. When participants used one of the data visualisation to make sense of the electricity consumption data, we told them that this was the feedback that Jack received from his Smart Meter In-Home Display.

The experiment involved three stages: the pre-test using the energy game, a period of exposure to energy usage visualisations, and a post-test using the energy game. In the game, participants made a series of 36 two-alternative forced-choices. The 36 comparisons crossed each of the nine appliances in the dataset. As described above, we recorded participants' response accuracy, response time, and decision confidence. In general, the energy used by the different appliances fell into several categories. The dishwasher, the washing machine and the vacuum cleaner were relatively high-energy consumption appliances. The light and the radio were relatively low-energy consumption appliances, with the microwave, the toaster, the coffee maker and the kettle being in between. This meant that some comparisons were relatively easy (e.g., dishwasher vs. light) and others were more difficult (e.g., dishwasher vs. washing machine). This range in difficulty meant that participants would have a range in decision accuracy and confidence. The focus here was to assess changes in participants' decisions between the different visualisation conditions. Participants received no feedback on the performance in the game (i.e. they would not receive feedback on whether their choices were correct or incorrect).

For the middle part of the experiment, participants saw a simulated pattern of appliance usage and were given feedback about the associated energy usage through the visualisation (Figure 24). The simulation had thirty frames, listed in Table 3. The simulation was designed to give periods in which different appliances were being used, sometime together, sometimes in isolation. The idea was to give a complex and rich pattern that mimicked domestic appliance use. Participants were free to look at each frame of the simulation for as long as they wanted to, proceeding through the experiment by clicking the continue button. For each given frame of the simulation, the nine household appliances were shown on the left side of the screen. These pictographs were the same that participants saw and clicked on in the energy game. Different combinations of appliances would be switched 'on' and 'off'. Figure 24 shows an example frame (Table 3, frame 11) in which the dishwasher is 'on', represented by a green background colour, while all other appliances are 'off'. On the right side of the screen, the data visualisation shows the associated energy usage for the appliances that are 'on'.

The frames of the simulation were the same between conditions. Which visualisation a participant saw on the right varied depending on which condition they were assigned to (Figure 20 - Figure 22).

Figure 24. Simulation.

Frame	Appliance(s) ON	Frame	Appliance(s) ON
1	Radio	16	Vacuum cleaner, microwave
2	Radio, lights	17	Radio
3	Radio, lights, kettle	18	Radio, toaster
4	Kettle	19	Radio, toaster, dishwasher
5	Kettle, toaster	20	Lights
6	Kettle, toaster, coffee maker	21	Lights, coffee maker
7	Coffee maker	22	Coffee maker, washing machine
8	Coffee maker, vacuum cleaner	23	Lights, microwave
9	Vacuum cleaner, dishwasher	24	Lights, microwave, toaster

10	Washing machine	25	Lights, microwave, toaster, coffee maker
11	Dishwasher	26	Microwave, radio
12	Dishwasher, washing machine	27	Microwave, lights
13	Washing machine	28	Kettle
14	Washing machine, vacuum cleaner	29	Kettle, vacuum cleaner, dishwasher
15	Vacuum cleaner	30	Vacuum cleaner, dishwasher, washing machine

Table 3. Summary over the thirty frames in the simulation.

Once they had finished the simulation, participants again completed the post-test energy game (which was exactly the same as the pre-test). After that, they were briefly interviewed to assess how they made sense of the visualisation in the simulation, and what they learned from it. They were given the opportunity to add any further comments.

5.5 Results

First, the quantitative results from the energy game are presented. For inferential statistical analysis, we use an Analysis of Variance (ANOVA) with a significance level of .05 for judging the significance of effects. Second, the qualitative data from the interview is presented.

5.5.1 Quantitative data

Figure 25 shows the results for response accuracy in the energy game (the proportion of correct decisions out of the 36 pairwise comparisons in percentages), Figure 26 shows response confidence (on a scale from 1 to 5, where 1 is low confidence and 5 is high confidence), and Figure 27 shows response time (in seconds). Tables 4, 5, and 6 list the corresponding descriptive means and standard deviations.

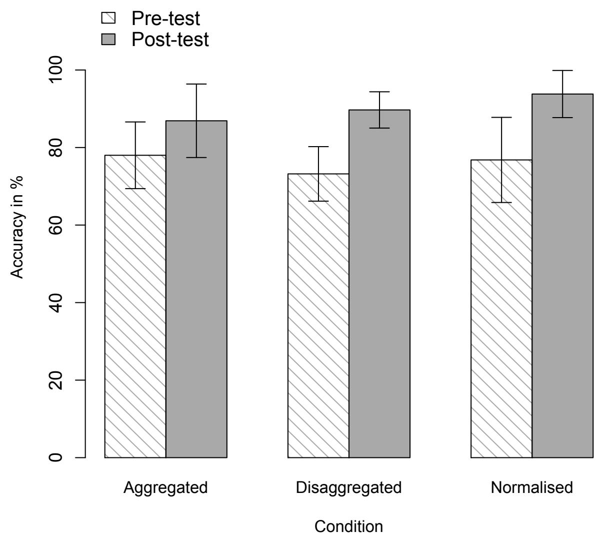


Figure 25. Response Accuracy.

	Aggregated	Disaggregated	Normalised	Total
Pre-test	77.96 (8.56)	73.21 (7.03)	76.79 (10.98)	76.03 (9.01)
Post-test	86.85 (9.48)	89.68 (4.68)	93.85 (6.08)	90.05 (7.53)

Table 4. Response Accuracy means and standard deviations, M(SD) in %.

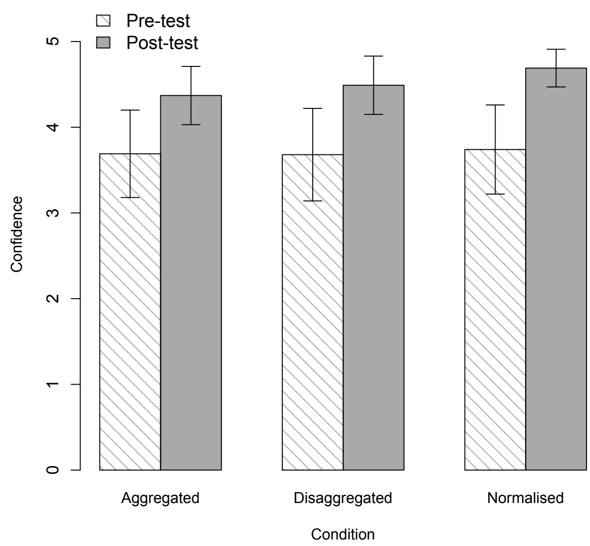


Figure 26. Response Confidence.

	Aggregated	Disaggregated	Normalised	Total
Pre-test	3.69 (.51)	3.67 (.54)	3.74 (.52)	3.7 (.51)
Post-test	4.37 (.34)	4.49 (.34)	4.69(.22)	4.52 (.33)

Table 5. Response Confidence means and standard deviations, M(SD).

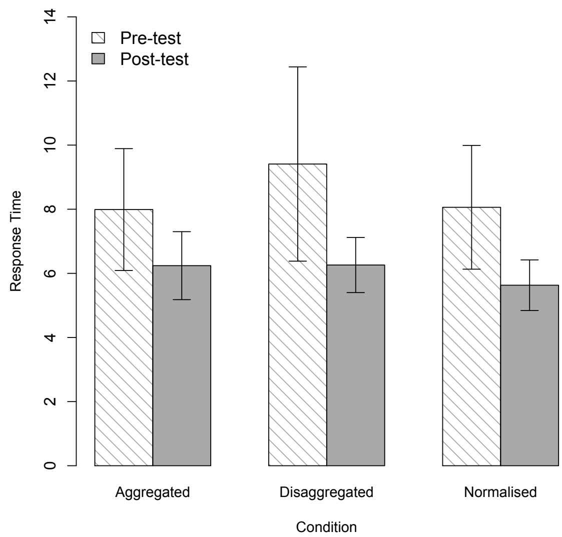


Figure 27. Response Time.

	Aggregated	Disaggregated	Normalised	Total
Pre-test	8 (1.9)	9.41 (3.03)	8.06 (1.93)	8.48 (2.37)
Post-test	6.24 (1.06)	6.26 (.86)	5.63 (.79)	6.05 (0.94)

Table 6. Response Time means and standard deviations, M(SD) in seconds.

The significant main effect of time of test (Table 7) shows that participants were better able to identify the appliance that used the most electricity at post-test than at pre-test. They were also more confident at post-test than at pre-test and they were faster at post-test than at pre-test. This main effect of time suggests that regardless of condition, participants improved their knowledge of how much electricity domestic appliances consume. The main effect of condition (i.e. visualisation) was not significant for any of the depended variables. Neither was there an interaction effect between condition and time of test.

	Main effect time	Main effect condition	Interaction effect
Accuracy	F(1,40)=76.10, p<.001**	F(2,40)=1.52, p=.23	F(2,40)=2.7; p=.08
Confidence	F(1,40)=123.12,p<.001**	F(2,40)=1.04, p=.36	F(2,40)=1.13, p=.33
Response time	F(1,40)=61.23, p<.001**	F(2,40)=1.77, p=.18	F(2,40)=1.65, p=.21

Table 7. Repeated Measures Anovas.

However, these tests average together pre- and post-test scores, so we might therefore not expect to see any effect. Given the retest nature of the experimental design, we therefore consider separately for pre-test or post-test scores whether there were effects of condition. There was no significant effect of condition on pre-test measures (Table 8). This confirms that all participants had a comparable level of background energy literacy and performed similarly on the pre-test regardless of which experimental group they were assigned to.

At post-test, after participants had been exposed to one of the three different data visualisations, there was a significant effect of condition on accuracy scores and on confidence scores. Table 9 shows the results of pairwise comparisons, using LSD adjustments to correct for making multiple comparisons. The pairwise comparisons found that participants had significantly better accuracy at post-test in the normalised condition than in the aggregated condition. Equally, participants in the normalised condition were significantly more confident in the post-test than participants in the aggregated condition.

	Pre-test	Post-test
Accuracy	F(2,40) = 1.08, p = .35	F(2,40) = 3.53, p = .04*
Confidence	F(2,40) = .06, p = .94	F(2,40) = 4.01, p = .03*
Response time	F(2,40) = 1.66, p = .2	F(2,40) = 2.12, p = .12

Table 8. Anovas comparing the three groups at pre- and post-test.

	Aggregated Disaggregated	Aggregated Normalised	Disaggregated Normalised
Accuracy	p = .29	p = .01**	p = .13
Confidence	p = .32	p = .01**	p = .08
Response time	p = .94	p = .08	p = .07

Table 9. Pairwise comparisons at post-test with LSD adjustment.

5.5.2 Qualitative data

Participants were asked how they had made sense of the data visualisation in the simulation. In the following, differences in how participants said they made sense of the data in the different conditions shall be explored.

Participants in the aggregated condition reported looking at how much the separate activities consume and how long they lasted for. When multiple devices were on at the same time, they tried to *'see how they add up'* (P4) and *'how much they consume all together minus individual ones'* (P3). To estimate the total consumption of one activity, they *'add[ed] up the energy they use in different periods'* (P4) in order to estimate the area under the curve. P10 stated that *'when they were combining, it made it more difficult to see and remember which one is more'*. Difficulties were reported with activities that were similar in the amount of electricity consumed, such as the coffee maker and the kettle. A couple of participants mentioned they were thinking about the particular patterns of the activities, such as the *'hot cycles'* of washing machine and dishwasher, which are mirrored in the *'the peaks and trough of the graph'* (P13).

Participants in the disaggregated condition reported *'looking at how the energy level changes. For comparable time, [I] look at the difference in height and kind of estimate the total area'* (P18). P25 found that *'Of course many things became clear (…) With the graphs you could estimate how much and the times when they consume. It was accurate in determining the pattern'* while P26 found it *'difficult to judge, there are* *all those spikes'*. Just as in the aggregated condition, the difficulty depended on how similar the activities were in the amount of electricity consumed: *'Some things were quite obvious like the radio, it's not consuming anything at all'* (P27), *'I was confused between laundry and dishwasher'* (P25). P29 described her memorising strategy as:

> *'trying to think of how it works, how the piece of technology works (…) I found [the graph of the dishwasher] interesting cause I thought it has two peaks and in the middle it is low so I was thinking okay so what does it do? It sprays water at the beginning; then in the low bit, does it mean that the dishes stay in soap? For whatever, 30 minutes. And then has another peak of rinsing. Maybe it's not true but that's the explanation that I gave myself'.*

Participants in condition three did not have to compare visually, as the visualisation provided the ranking by consumption. They reported their strategy as *'see the curve and try to remember the sequence'* (P37), particularly trying to remember *'which ones took less (…) when there where small differences'* (P33). P40 thought *'the curves were pretty transparent, it was easy to see which one was higher (…) with kettle, lamp, coffee maker and toaster it was easy, they were one above another'*. On the other hand, we had to exclude P31 from the quantitative data analysis because she reported that the graph *'didn't make sense'* and she was unclear *'what the whole thing, the curvy shape was'* and admitted she had just clicked through the experiment. P44 was unsure *'if they [the graphs] were cumulative. I think they were not cumulative'* and would have liked to see the pattern that the appliances produce over time, yet he *'liked they were standardised over time, that was nice'.*

5.6 Discussion

5.6.1 Main findings

It is assumed that disaggregated feedback is more useful for householders to understand how much energy everyday practices consume (Darby, 2001; Fischer, 2008; Froehlich et al., 2011). We expected participants in the disaggregated group to outperform participants in the aggregated group, because disaggregated line graphs offer a level of information that is obscured in aggregated line graphs (Loorak et al., 2016). There were no statistical differences between the aggregated and disaggregated conditions. Hence, we cannot confirm the assumption that disaggregated feedback is per se more suitable than aggregated feedback. This might be due to limitations of the experiment (see 5.6.3 for details), or it might indicate that disaggregation alone visualised as time-series data is not sufficient.

The other assumption was that participants in the normalised group would outperform both other groups, because area-based graphs are more suitable than line graphs to summarise consumption over time (Costanza et al., 2012). Indeed, the statistical tests showed that the normalised group achieved significantly higher post-test scores than the aggregated condition. This finding confirms that data comprehension depends on the manner of presentation and design (Chiang et al., 2012; Roberts & Baker, 2003; Yun et al., 2010) and it suggests that disaggregation may be necessary, but not sufficient.

This is tentative evidence that summarised visualisations that de-emphasise time are more suitable for people to learn how much energy practices consume. An energy-centric visualisation, as opposed to a time-centric visualisation showing power fluctuations, might be easier to learn from because it is conceptually in line with people's mental models (how much energy is consumed), even if it is not representing structurally what cumulative energy usage is (power over time) (Cheng & Barone, 2017; Pinker, 1990; Zhang & Norman, 1994).

5.6.2 Reflection

The qualitative data yielded insights into the cognitive sense-making processes that participants went through in the different conditions. The aggregated condition was difficult to decode which is in line with the quantitative results. However, participants sometimes considered the cycles of the appliances with the ups and downs. It is possible that the cognitive effort leads to deeper processing which might be relevant for long-term retention. In the disaggregated condition, some information was immediately visible and easy enough to learn and remember (e.g. the radio consuming very little).

One participant spontaneously described ‘something interesting’ in the graph (Peebles et al., 2013). She shared her reflection on the dishwashers’ power consumption pattern, reasoning about what the dishwasher is doing in different stages of its cycle. Her comment is similar to the theme we found in Study 4, where participants were reading *beyond* the data. It could also be interpreted as an extrapolation from the data to the functioning of the device (Galesic & Garcia-Retamero, 2011). Beyond shining a light on the cognitive process of understanding the graph, the qualitative responses could also be analysed as an indicator of participants’ energy literacy – they may reveal how much someone knows about the energy consumption and processes involved in using appliances.

The normalisation is stripped of the characteristic consumption patterns, which one participant was curious to see. The normalised visualisation did not trigger any deeper reflection, learning was almost like memorising a list view (which is easy, but potentially defies the purpose of using a visualisation in the first place). It is also worth noting that one participant had to be excluded from the analysis because she did not understand the normalised visualisation and just clicked through the experiment.

The qualitative findings challenge the quantitative findings in that the normalised visualisation is not well suited, all things considered. Even though we collected very little qualitative data, there is some evidence that time series data triggers more reflection (as was expected).

5.6.3 Limitations

The biggest limitation to the validity of the findings are the high scores that participants achieved across conditions. Compared to the energy literacy measures from Study 2 and Study 3, the energy game is a playful approach that avoids technical questions and numeracy skills. Like a ranking task (which we used in Study 2), the energy game can generate a profile of correct and incorrect responses per participant (instead of asking them to name only one particularly high consuming device, as we did in Study 3). However, it is possible that the energy game was overall too easy to reveal differences between the groups, who might have achieved ceiling effects in their scores across the three conditions. All three groups learned significantly from pre-test to post-test, even if the learning increase was the smallest in the aggregated condition. Surprisingly, the disaggregated group was not statistically better than the aggregated group, and neither was the normalised group better than the disaggregated group. This might indicate that our design was too easy, i.e. the comparisons in the energy game were too easy or maybe it was not as difficult as we thought to learn the relevant information from the line graphs. The simulation showed isolated cycles of appliances at high resolution, which made it relatively easy to learn even in the most difficult condition.

The second limitation is that out of 43 participants, 36 were students (under-graduates and post-graduates). This means that the sample was relatively young, and probably more highly educated with better computer literacy than the general population. Moreover, the majority of the sample was female. Locoro, Cabitza, Actis-Grosso, and Batini (2017) found that the ability to understand infographics might be subject to age, gender, and educational background. Our sample's demographics imply that caution should be exercised in generalising our findings to the general population (Sturm et al., 2015). To some extent, the high level of education may explain the overall high scores in the energy game.

The third limitation is that participants saw someone else's data in a lab setting, which is certainly less meaningful than reflecting on one's own data in real life. It is unclear to what extent findings from laboratory experiments can transfer to uncontrolled settings in the real world (Rogers, Yuill, & Marshall, 2013). For example, Chiang, Natarajan, and Walker (2012) replicated their laboratory energy display study in the field and found slightly different results (Chiang, Mevlevioglu, Natarajan, Padget, & Walker, 2014). Even though experiments do not capture human perception in a real-world setting, they do provide 'a useful upper-bound on people's ability' (Chiang et al., 2012) and simulations are an appropriate and rigorous method to test cognitive abilities free from confounding variables (Gonzalez, Thomas, & Vanyukov, 2005).

The fourth limitation is that our operationalisation was one very specific task (i.e., making decisions in the energy game). We focused on one aspect of energy literacy, that of participants' learning about how much energy the appliances consume. In a real-world setting, people might be interested in learning various other things about their domestic energy consumption, e.g., which appliances are inefficient and eligible for retrofitting, at what time of day they are using most energy if they are in a time of use tariff, or what is contributing towards their baseline consumption (Van Dam, Bakker, & Van Hal, 2012). Our method was focused on energy-centric learning and neither the energy game nor our open-ended questions created sufficient opportunity to demonstrate time-centric learnings that were favoured by the line graph conditions. It needs to be further investigated how time-sensitive information (e.g., time-of-use tariffs) are best communicated to the user. Feedback showing trends over the day (like line graphs do) will be indispensable for that kind of scenario. Possible measures for data comprehension could be tasks that require participants to determine the best time of day to run certain household appliances. Ideally, they would combine the knowledge of which appliances consume most energy (appliance-centric) and when it would be sensible to run them (time-centric).

5.7 Conclusion

Study 5.1 found some evidence that disaggregation alone does not make energy feedback more useful. Only when presented in an energy-centric manner did disaggregated information increase participants understanding of how much energy practices consume. The disaggregated line graph was not any better than the aggregated line graph, which indicates that time-centric visualisations are not suitable for people to learn how much energy practices consume. Yet, time-centric data seemed to trigger deeper reflection in participants.

5.8 Introduction to Study 5.2

The first experiment normalised appliances' power consumption pattern over time to create an energy-centric visualisation. On the one hand, participants in the energy-centric visualisation had higher post-test accuracy than participants in the time-centric line graphs. On the other hand, the approach to creating an area-centric visualisation needs improving, because normalisations are unfamiliar to people and they are better at interpreting familiar diagrams (Peebles et al., 2013). Study 5.2 is an extension of Study 5.1, examining two new conditions, i.e. additional participants were recruited to test bar charts and bubble charts to see if they are superior to the normalised charts from Study 5.1.

Bar charts might seem an obvious choice, together with line graphs they are the most commonly used charts and people are familiar with them. The elementary perceptual task in decoding the information is simple, as it only involves a judgement of line length (height of the bar), and with all bars starting from the same origin (the x-axis), comparing between bars is easy (Cleveland & McGill, 1984). However, as opposed to the normalised visualisation from Study 5.1, they are not a true area-chart. As just explained, the one meaningful dimension of a bar chart is its length. The width of the bar is arbitrary and does not carry information. It is therefore different from all three conditions in Study 5.1, where even the line graphs created an area under the curve which reflected the amount of energy consumed. As energy is the product of power over time, the metaphor of an area-based chart to represent the amount of energy consumed, seems fitting.

Figure 28 shows an example of disaggregated energy data feedback using bubble charts. Bubble charts, as opposed to bar charts, are area-based, representing the amount of energy consumed by their size. The drawback is that the perceptual task is less clear and straightforward, as one might estimate the area or maybe focus on the diameter of a bubble. Particularly when comparing between bubbles this adds difficulty, since they do not share a common origin (Cleveland & McGill, 1984). Also, they are less familiar to people than bar charts and line charts (or even pie charts). For example, Murugesan, Hoda, and Salcic (2015) reviewed 22 studies that investigated energy data visualisations. Among them were line graphs, bar charts, box charts, pie charts, spiral displays, time charts, as well as abstract and artistic representations, but no bubble charts. This might indicate a mismatch between academic studies testing prototypes in one-off studies, and commercial products that tend to use conventional bar, pie, or bubble charts (Figure 18, Figure 19, Figure 28).

Figure 28. Disaggregated bubble chart.

The purpose of Study 5.2 is to test two energy-centric visualisations that are more familiar to people than the normalised visualisation from Study 5.1. The experimental setup in Study 5.2 is an exact replication of Study 5.1, extending the design to five between-subjects conditions. Study 5.2 does not collect new data for the three conditions from Study 5.1. Study 5.2 collects new data for the bar and bubbles charts and compares the results to the data collected in Study 5.1. The hypothesis is that the bar and bubble charts will outperform the time series charts and the normalised visualisation from Study 5.1.

5.9 Method

5.9.1 Sample

Thirty-five participants (12 male) were recruited through the UCL Psychology Subject Pool. Eight participants were aged between 18 and 20 years, 24 were between 21 and 35 years, and three were 36 years or older. All were adults with normal or corrected to normal vision who were accustomed to reading from left to right and who pay their utility bills (or do so with the help of their partners or fellow tenants). Participants received course credit or a small payment for taking part in the study.

5.9.2 Materials

Study 5.2 is a replication of Study 5.1. The underlying data set, the simulation, and the energy game remained the same and the experiment was run on the same computer in the same lab (5.4.2). Two new data visualisations were tested, the bar charts and the bubble charts (Figure 29). Both visualisations show energy consumed, the bubbles by area size, the bars by their y-value.

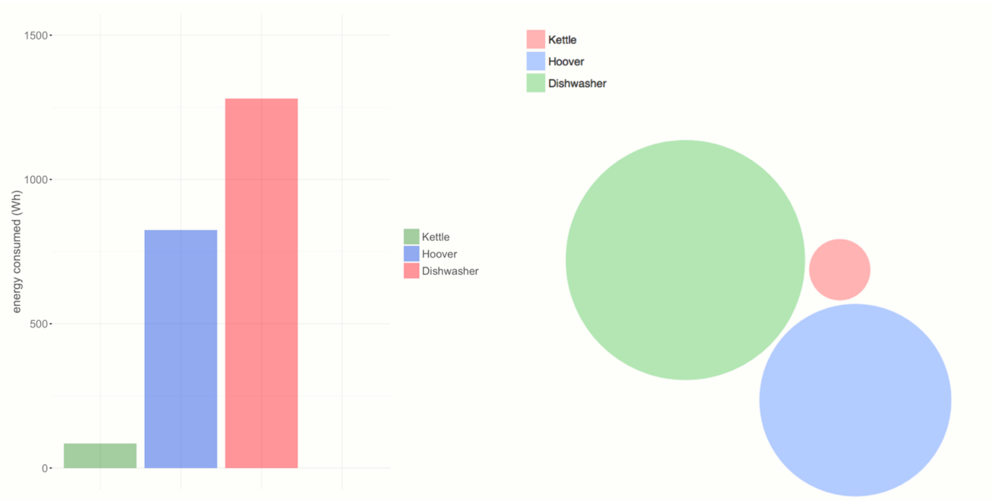


Figure 29. Study 5.2, bar and bubble visualisations.

5.9.3 Design

Study 5.2 is a between-subjects design with the visualisation as the independent variable and the performance in the energy game as the dependent measure.

5.9.4 Procedure

The procedure is a replication of Study 5.1 (5.4.4) with two conditions, the bar and bubble charts.

5.10 Results

5.10.1 Quantitative data

The quantitative analysis combines the data set from Study 5.1 with the data collected in Study 5.2. Figure 30 and Table 10 show the means and standard deviations for the five groups in terms of their response accuracy. Like the first three groups, the bar group and the bubble group increased their accuracy scores from pre-test to post-test. As can be seen from the descriptive statistics, the two new conditions, the bars and the bubbles, yielded lower accuracy scores in the post-test than the normalised visualisation. The Analysis of Variance at post-test and the pairwise comparisons confirm that the bar and the bubble group were not statistically different from any of the three groups from Study 5.1 (Table 14 and Table 15).

These statistics already refute our hypothesis that the bars or bubbles would achieve better post-test scores than the normalised visualisation. Accuracy in the energy game is the main dependent variable of interest, because it shows if and how much participants have learned. Confidence and response time (Table 11 and Table 12) are less meaningful if there are no differences in accuracy. However, participants in the bar and bubble condition show both a main effect of time of test and of condition (Table 13), as they were responding faster both at pre-test and at post-test (Table 14) than the aggregated and the disaggregated condition (Table 15). Participants in the normalised condition, in the bar condition, and in the bubble condition, were also more confident at post-test than participants in the aggregated condition (Table 14 and Table 15).

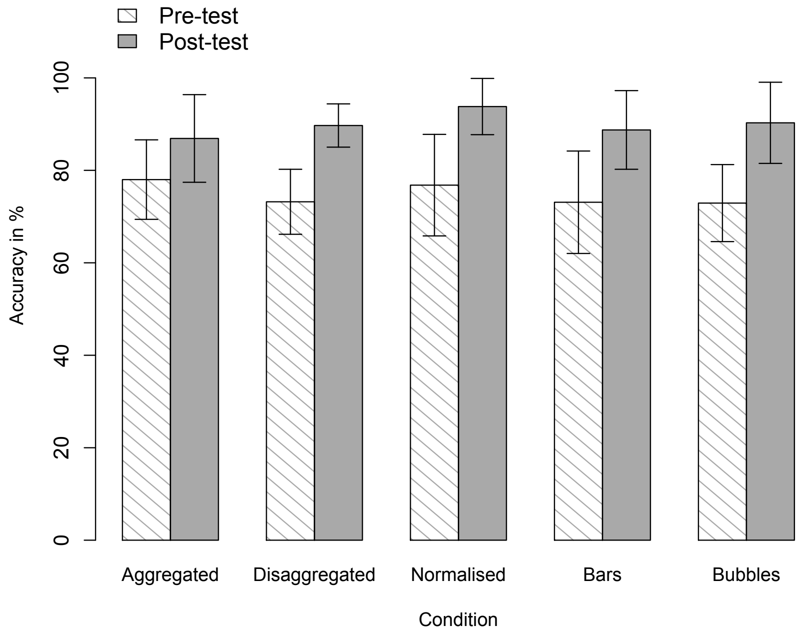


Figure 30. Response Accuracy.

	Aggregated	Disagg.	Normalised	Bars	Bubbles
Pre-test	77.96 (8.56)	73.21 (7.03)	76.79 (10.98)	73.1 (11.08)	72.92 (8.33)
Post-test	86.85 (9.48)	89.68 (4.68)	93.85 (6.08)	88.74 (8.51)	90.28 (8.54)

Table 10. Response Accuracy means and standard deviations, M(SD) in %.

	Aggregated	Disaggregated	Normalised	Bars	Bubbles
Pre-test	3.69 (.51)	3.67 (.54)	3.74 (.52)	3.98 (.57)	3.83 (.48)
Post-test	4.37 (.34)	4.49 (.34)	4.69(.22)	4.63 (.29)	4.61 (.38)

Table 11. Response Confidence means and standard deviations, M(SD).

	Aggregated	Disaggregated	Normalised	Bars	Bubbles
Pre-test	8 (1.9)	9.41 (3.03)	8.06 (1.93)	7.15 (1.78)	7.46 (1.32)
Post-test	6.24 (1.06)	6.26 (.86)	5.63 (.79)	5.2 (.95)	5.3 (1.03)

Table 12. Response Time means and standard deviations, M(SD) in seconds.

	Main effect time	Main effect condition	Interaction effect
Accuracy	F(1,73)=115.57, p<.001**	F(4,73)=1.22, p=.31	F(4,73)= 1.23, p = .31
Confidence	F(1,73)=190.08, p<.001**	F(4,73)=1.56, p=.19	F(4,73)=.89, p=.48
Response time	F(1,73)=124.67, p<.001**	F(4,73)=3.94, p=.01**	F(4,73)=1.33, p=.27

Table 13. Repeated Measures Anovas.

	Pre-test	Post-test
Accuracy	F(4,73) = .99, p = .42	F(4,73) = 1.57, p = .19
Confidence	F(4,73) = .96, p = .44	F(4,73) = 2.4, p = .05*
Response time	F(4,73) = 2.8, p = .03*	F(4,73) = 4.47, p = .01**

Table 14. Anovas comparing the three groups at pre- and post-test.

Comparison		Accuracy	Confidence	Response Time
Aggregated	Disaggregated	p = .33	p = .34	p = .94
	Normalised	p = .02*	p = .01**	p = .09
	Bars	p = .49	p = .02*	p = .01**
	Bubbles	p = .23	p = .04*	p = .01**
Disaggregated	Normalised	p = .16	p = .09	p = .08
	Bars	p = .73	p = .21	p = .01**
	Bubbles	p = .84	p = .3	p = .01**
Normalised	Bars	p = .07	p = .58	p = .21
	Bubbles	p = .22	p = .48	p = .35
Bars	Bubbles	p = .56	p = .85	p = .76

Table 15. Pairwise comparisons at post-test with LSD adjustment.

5.10.2 Qualitative data

Participants were asked how they made sense of the data visualisations. They said they liked the bar charts and thought the visualisation was *'nice'*, *'helpful'* and *'useful'*. They reported learnings, for example, one participant said before the simulation he thought the radio would consume more energy than lights. Participants would compare appliances and *'try to remember how much they consume compared to each other'*. Participants in the group with the bubble charts, too, reported comparing, ordering and memorising the relative size of the circles. One participant found it hard to differentiate between the vacuum cleaner and the washing machine and between the kettle and the coffee machine because the bubbles are very similar. Another participant found the bubbles *'off-putting'*. Others reported learnings similar to the bar chart group, such as finding out that the vacuum cleaner was comparable to the washing machine and that the radio consumes surprisingly little.

5.11 Discussion

5.11.1 Main findings

Across Study 5.1 and Study 5.2, all groups increased their accuracy scores from pre- to post-test. The only group that performs significantly better in the post-test is the normalised group compared to the aggregated group. The motivation of Study 5.2 was to improve the approach to an energy-centric visualisation and to replace the normalised graph with more familiar charts, i.e. the bars and bubbles. It appears that this did not work.

Participants in the bubble condition reported difficulties in comparing the circles when they are similar in size. This was to be expected based on graph theory, because the comparison of graphical elements is harder when they do not share the same origin (like multiple bars in a coordinate system do) (Cleveland & McGill, 1984). The bubbles convey information on how much energy was consumed through their size (area), but it is unclear how people judge the size, i.e. which element of the chart they are focusing on (total size, or simple diameter, or even circumference).

As for the bars, they show energy-centric information like the normalised visualisation and the bubbles, but they are different in that they are not area-based charts. As explained in the Introduction, they only code information through the bars' length, the width-dimension of bars is meaningless. That means, in judging one bar, the elementary perceptual task is to assess the height of the bar (vertical). The task of comparing different bars involves two steps, first assessing the height of every bar (vertical) and then to compare their heights across bars (horizontal). In comparison, the normalised visualisation involves vertical judgements only: the visual tasks are to judge and compare the height of the curves, which are both in the vertical dimension. This is a hypothesis based on the data available from the two presented experiments, which would need thorough investigation in cognitive experiments examining this in more detail, potentially using eye-tracking to determine people's gaze patterns.

5.11.2 Reflection

The qualitative data collected in Study 5.2 was similar to Study 5.1. Like the normalised graph, the bars and bubbles did not trigger deeper reflection that went beyond the data. One participant expressed his dislike for the bubble charts.

5.11.3 Limitations

We collected the data for the bar and bubble charts as a follow-up to the first experiment. We then compared the data to the data collected in the first experiment. This is a weakness in the data collection and analysis. For unknown reasons, there might be differences between the two samples, and indeed we found response time differences at pre-test. A large part of the sample consists of university students; different times of data collection might imply that students in one sample might have been more attentive or less focused (e.g. due to ongoing exam periods). The bar and bubble graphs were hypothesised to be the easiest condition, yet they did not yield the highest accuracy scores in the post-test. The finding that an unfamiliar normalised graph yield better results than a bar or bubble chart has little face validity. Study 5.2 challenges the conclusions from Study 5.1. Considering both experiments' findings, there is no conclusive evidence that energy-centric charts are superior to line graphs for communicating appliance-centric energy consumption.

The limitations discussed in Study 5.1 (5.6.3) still hold for Study 5.2. That us, there is a lack of generalisability due to a mostly female student sample and limited ecological validity. Also, as discussed in Study 5.1, the energy game might have been too easy, judging from the overall very high scores that participants achieved.

5.12 Conclusion

The results from Study 5.2 indicate that bar and bubble charts were not any more effective than time series graphs. Across the five visualisations tested in Study 5.1 and 5.2, the normalised visualisation yielded the highest post-test accuracy scores in the energy game.

5.13 Introduction Study 5.3

Studies 5.1 and 5.2 indicate that the normalised visualisation was the most effective condition for participants to learn how much energy practices consume. Surprisingly, the disaggregated time series was not more effective than the aggregated time series visualisation. The first question is whether there really is no difference between the two, or if it wasn't detected. The second question is how to represent energy-centric information other than in a normalised visualisation.

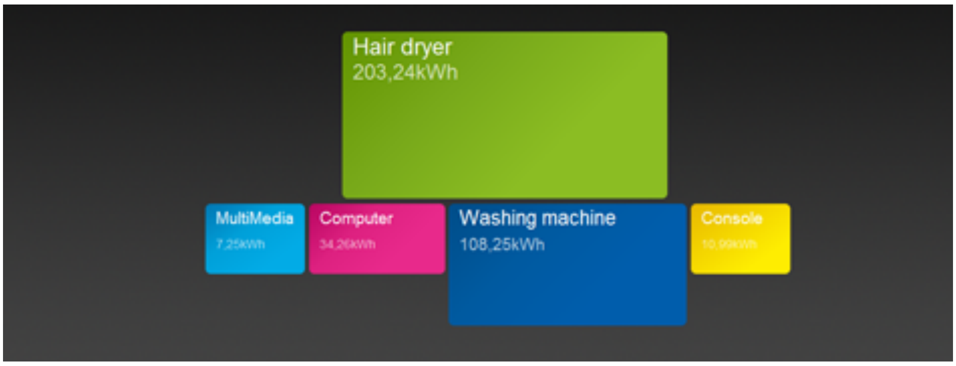
To address these questions, Study 5.3 replicated the previous experiments with several modifications (specified in the following and in the Method 5.14). It again tests 1) an aggregated time series line graph, 2) a disaggregated time series line graph, and 3) an area-based energy-centric visualisation.

To investigate the first question posed above (i.e. are disaggregated line graphs no more efficient than an aggregated line graph?), Study 5.3 addresses a couple of shortcomings from Study 5.1. The first experiment showed isolated runs of appliances, which was perhaps too simplistic. It is possible that even in the aggregated condition, it was relatively easy to determine and memorise the relative share of an appliance. Instead of showing isolated runs, Study 5.3 shows energy consumption patterns for 24 hours in the simulation. Further, we added two appliances, a fridge and a TV, to the nine appliances that were used in the first two experiments (which where: radio, lamp, microwave, toaster, kettle, coffeemaker, vacuum cleaner, washing machine, and dishwasher) to make the simulation more complex and to add a baseline of always-on consumption (i.e. the fridge). Findings from Studies 3 and 4 were that people neglect the baseline in time series line graphs, and focus on peaks (which is what the visual system is more attuned to; Tufte, 1983).

The second question is how to represent energy-centric information other than in a normalised visualisation. A seemingly straightforward approach to transforming a disaggregated time series graph into an energy-centric area-based visualisation would be to maintain the time information and to eliminate the power fluctuations by using the average power as a constant y-value. This would create a two-dimensional rectangular shape. The problem with this approach is that it does not work in practice. For example, plotting a kettle (2 minutes|2,000W), a radio (60 minutes|5W), and a fridge (24 hours|100W) in the same coordinate system would hardly be legible, as it would result in a confusing visual disarray with boxes that vary greatly in height and width and some appliances would barely be visible. Therefore, one more transformation is needed to reshape these rectangles into squares, or at least rectangles with sides that vary less in length.

The approach of visualising energy-centric information in form of rectangles has been tested by FigureEnergy's Consumption Overview (4.13.2, Costanza et al., 2012). Unfortunately, participants did not engage with the Consumption Overview much (neither in the first FigureEnergy study in 2012, nor in Study 4 presented in this thesis). However, there are other publications that deployed and tested box-shaped representations.

Schwartz, Denef, Stevens, Ramirez, and Wulf (2013) conducted a longitudinal study, collecting both total and disaggregated data through smart plugs. They found that people reflected on and learned from the appliance-level feedback, comparing the energy consumption of different appliances to each other. Figure 31 shows how they visualised appliance-level data in two-dimensional blocks.



Comparative Tag Cloud: The tag cloud shows sums of consumption of *SmartPlugs* grouped by user-generated tags.

Figure 31. Appliance-level visualisation by Schwartz et al. (2013).

Borghouts, Soboczenski, Cairns, and Brumby (2015) used two-dimensional blocks to represent the magnitude of numbers, where the blocks' width encoded the order of magnitude of a number, and the height the overall value (Figure 32). This representation was successful in helping participants spot numeric entry errors.

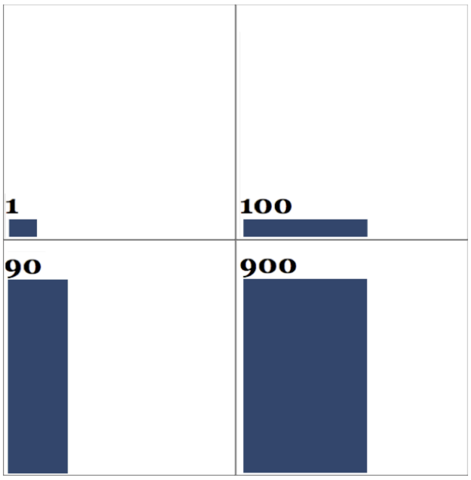


Figure 32. Block representations by Borghouts et al. (2015).

Using rectangle shapes offers another convenient opportunity, namely that of superimposing pictographs (pictorial symbols, or icons, representing a word or an item) onto the representation, like Costanza, Ramchurn, and Jennings (2012) did in the Consumption Overview of FigureEnergy, and like commercial providers do (Figure 28). Pictographic representations have been found to be suitable for quick cognitive encoding and retention (Haroz, Kosara, & Franconeri, 2015), and they might help to make the information visualisation more engaging and aesthetically appealing, which is a criterion that makes energy feedback more effective (Fischer, 2008).

Finally, the last modification we made to the experimental procedure was to add questions to the qualitative interview at the end. The responses in the previous experiments did not generate a lot of data, hence, we aimed to explore in more depth how participants read beyond the data. Therefore, we asked them how they thought Jack (the persona used in the experiment) could reduce his energy consumption, and whether they could explain why the appliances use the amount of energy they consume. These questions were inspired by the reflections of participants in the previous disaggregated condition in Study 5.1 (e.g. their reasoning about how the dishwasher consumes energy). The aim of these questions was to see how deeply participants reflected, if the time-centric visualisations triggered more reflection, and if we could gain deeper insights into participants' graphical and energy literacy.

5.14 Method

5.14.1 Sample

Sixty-eight participants (41 female) were recruited through the UCL Psychology Subject Pool. Three participants were aged between 18 and 20 years, 57 were between 21 and 35 years, and five were 36 years or older. All were adults with normal or corrected to normal vision who were accustomed to reading from left to right and who pay their utility bills (or do so with the help of their partners or fellow tenants). Participants received course credit or a small payment for taking part in the study.

5.14.2 Design

Study 5.3 is a between-subjects design with the visualisation as the independent variable and the performance in the energy game as the dependent measure. The three visualisations are 1) an aggregated line graph, 2) disaggregated line graphs and 3) an area-based energy-centric visualisation, further described in Materials (5.14.3).

5.14.3 Materials

Study 5.3 is a replication of Study 5.1 and 5.2. The underlying data set (UK-DALE), the simulation, and the energy game remained the same and the experiment was run on the same computer in the same lab (5.4.2).

The following changes have been made in comparison to the previous two experiments:

The fridge and the TV have been added to the previously tested nine appliances (radio, lamp, microwave, toaster, kettle, coffeemaker, vacuum cleaner, washing machine, and dishwasher). As explained in the introduction, this was done to create a richer data set and to provide a baseline (the fridge). The data for the fridge and the TV were taken from the same dataset (the UK-DALE dataset (Kelly & Knottenbelt, 2015)). This leads to the energy game (Figure 23) increasing from the 35 two-alternative forced-choice decisions to 55 pairwise comparisons to cover each combination of the eleven appliances.

The simulation has been changed from showing 30 frames with isolated runs of appliances, to seven frames showing energy profiles of 24-hour periods at a time, one frame per day of the week (Figure 33, Figure 34, Figure 35). Again, this was done to create a richer data pattern, to make the data feedback more realistic, and to allow for more variance in learning, seeing that we found ceiling effects for participants' performance in the energy game in Study 5.1 and 5.2 where we used isolated runs of appliances. The seven frames were meant to imitate a typical week in Jack's house (the persona), where there was more activity during the weekend than on weekdays. Sometimes appliances were used in isolation, sometimes they would overlap. There were appliances that would consume a lot of electricity for a few minutes only (e.g., the kettle), and others that would consume less power over a longer period (e.g., the TV). The fridge generated a constant 24-hour baseload.

The visualisations tested in Study 5.3 are depicted in Figure 33, Figure 34, and Figure 35. The aggregated condition (Figure 33) shows the total energy consumption, merely labelling the onsets in time when the appliances were turned on (like the

Consumption Graph in FigureEnergy does). This was done to provide participants with an understanding of the pattern of domestic activities that were taking place in the home. Some information is visible enough in this condition: one can detect the baseline of the fridge, for example, and the spike of the kettle is distinct. On the other hand, it is nearly impossible to make out the exact pattern of the dishwasher.

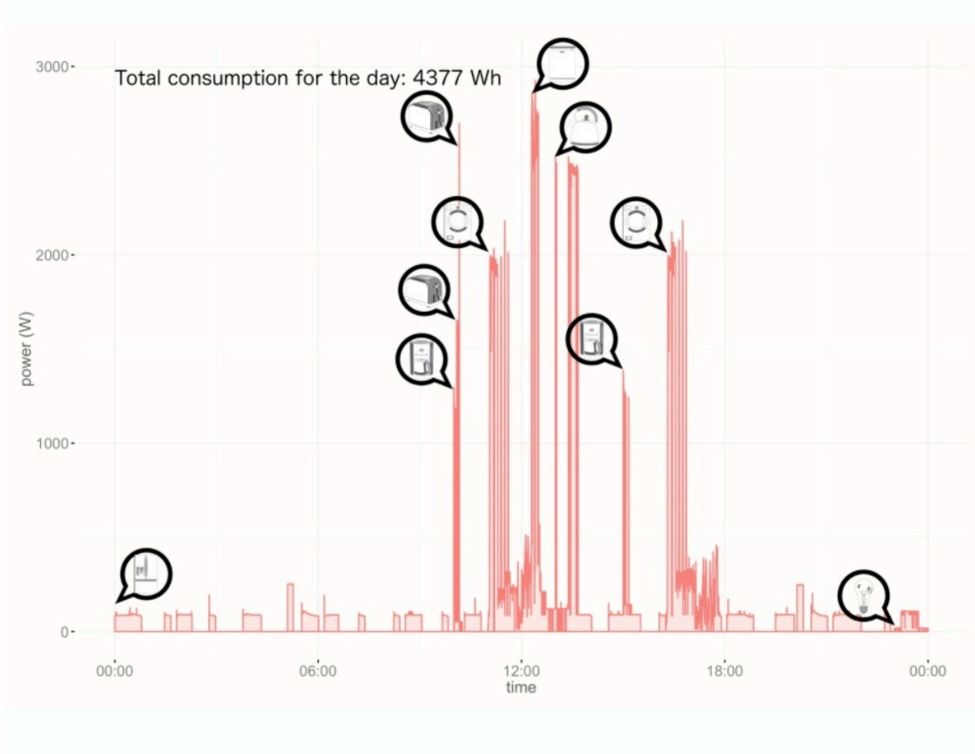


Figure 33. Study 5.3, aggregated visualisation.

In the disaggregated condition (Figure 34), the dishwasher's distinct pattern with two peaks and a period of lower usage in the middle becomes visible. Another example that exemplifies the difference between the aggregated and disaggregated visualisation is the second usage of the toaster – in the aggregated condition the toaster and coffee maker add up to a higher peak and the contribution of each becomes blurred. In this way, it seems intuitive that the disaggregated data visualisation should aid participants as they make sense of how much electricity different devices are consuming.

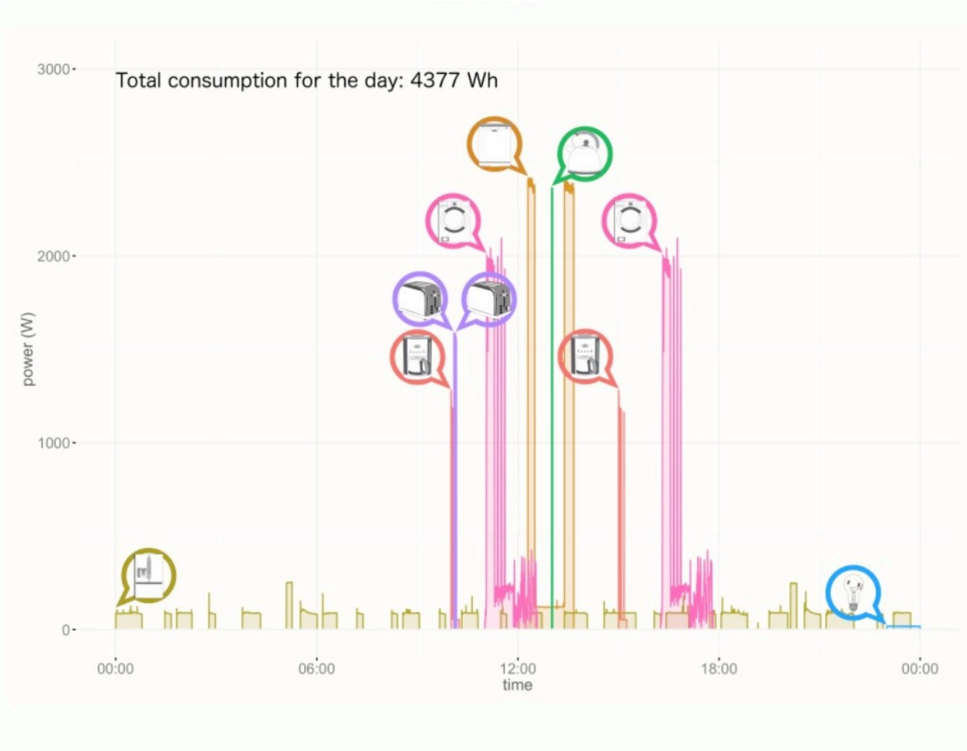


Figure 34. Study 5.3, disaggregated visualisation.

In the third condition (Figure 35), participants were presented with an area-based energy-centric visualisation using pictographs (this condition will be referred to as pictographic). Every pictograph showed the cumulative energy consumed by an appliance during the day. The size of the pictograph was proportional to the energy consumed by the appliance. For example, if one appliance consumed twice as much energy as another, the pictograph of one will be twice as big in area as the other. If 'Jack' (the persona) used the same appliance twice in a day, then the visualisation will show two pictographic blocks for the appliance.

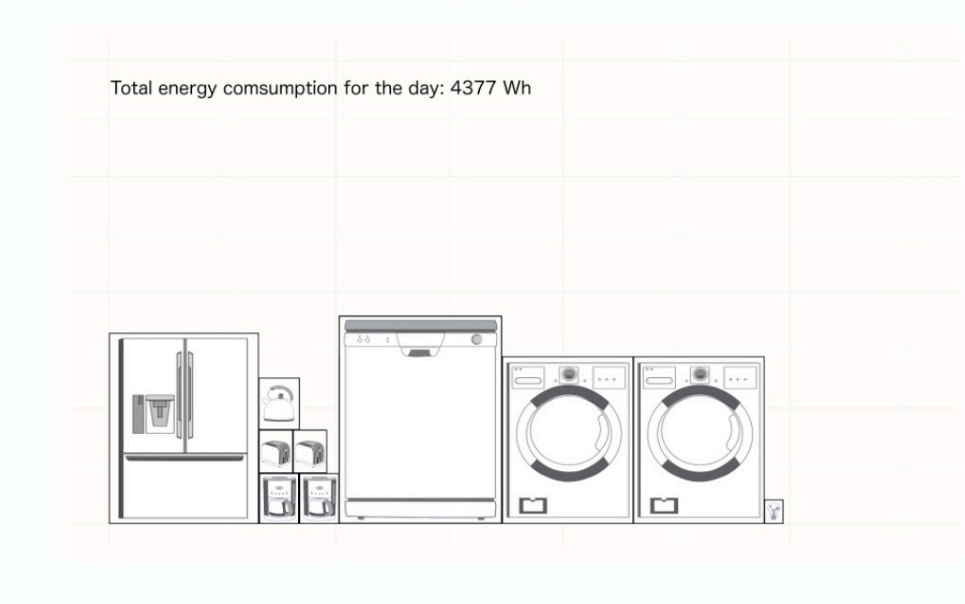


Figure 35. Study 5.3, pictographic visualisation.

The changes we made to the interview at the end was to add the following two questions: ‘How could Jack reduce his consumption?’ and ‘Can you explain why the appliances use the amount of energy they consume?’.

5.14.4 Procedure

The procedure is a replication of Study 5.1 and 5.2 (5.4.4) with the changes just described in Material (5.14.3).

5.15 Results

5.15.1 Quantitative data

Three participants admitted to having clicked through the simulation without trying to learn from it and were therefore excluded from the quantitative data analysis, reducing the sample size from 68 to 65.

Figure 36 and Table 16 show the descriptive results for response accuracy in the energy game (the proportion of correct decisions out of the 55 pairwise comparisons in percentages). Table 19 shows that there was no main effect of assigned condition on participants’ accuracy. As in Study 5.1, we would not expect to see a main effect of condition because this analysis averages together pre-test and post-test scores. As expected, there was a significant main effect of time of test (Table 19). Figure 36 illustrates that the main effect of time of test is driven by the pictographic group who increased their score from pre-test to post-test by almost 18% (Table 16). Anovas comparing the three groups at pre-test and at post-test show that the groups had comparable scores before being exposed to the data visualisations, but at post-test, the groups' scores were significantly different (Table 20). Post-hoc comparisons with LSD adjustments to correct for making multiple comparisons, confirm that participants had significantly better accuracy at post-test in the pictographic condition compared to both the aggregated and the disaggregated condition (Table 21). There was a significant interaction effect of time of test and condition, confirming that the learning curve between the groups was different (Table 19).

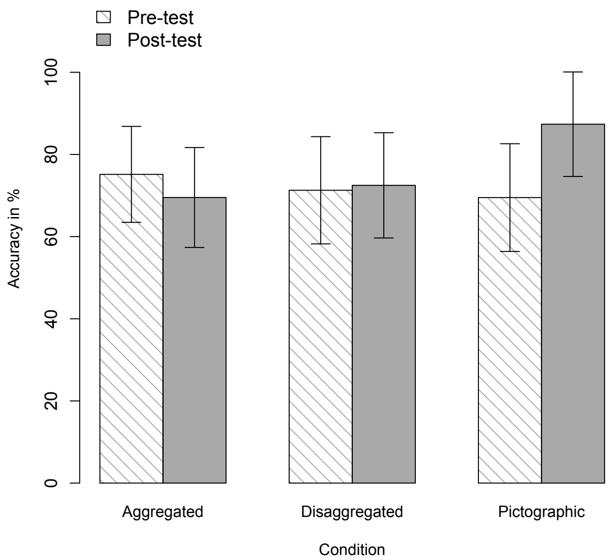


Figure 36. Response Accuracy.

	Aggregated	Disaggregated	Pictographic	Total
Pre-test	75.15 (11.68)	71.82 (13.05)	69.5 (13.11)	72.11 (12.66)
Post-test	69.52 (12.16)	72.48 (12.81)	87.36 (12.72)	75.56 (14.67)

Table 16. Response Accuracy means and standard deviations, M(SD) in %.

Figure 37 and Table 17 show the descriptive results for response confidence (on a scale from 1 to 5, where 1 is low confidence and 5 is high confidence). The main effect of time of test is significant, showing that participants in all conditions became more confident from pre-test to post-test, while there was no main effect of condition and no significant interaction effect (Table 19). Neither were there any differences between the groups at post-test (Table 20 and Table 21).

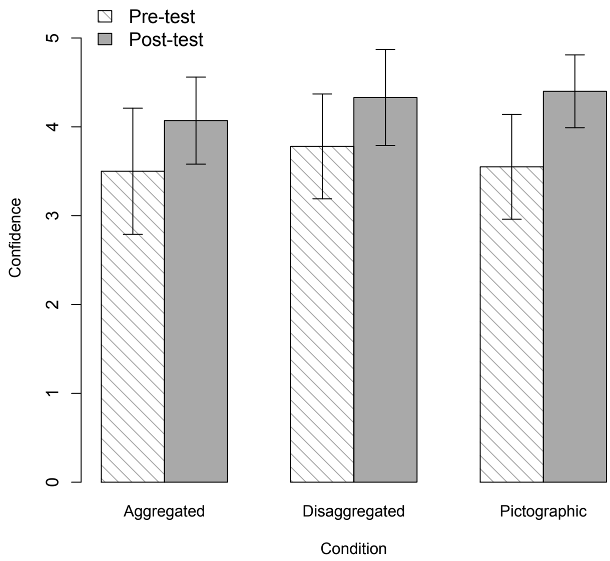


Figure 37. Response Confidence.

	Aggregated	Disaggregated	Pictographic	Total
Pre-test	3.64 (.83)	4.09 (.63)	3.86 (.62)	3.87 (.71)
Post-test	3.93 (.43)	4.01 (.63)	4.09(.69)	4.01 (.59)

Table 17. Response Confidence means and standard deviations, M(SD).

Figure 38 and Table 18 show response time (in seconds). The descriptive statistics for response time show that the participants seemed to become faster at giving a response from pre-test to post-test, but there was no main effect of time of test. Neither was there an effect of condition, or an interaction effect (Table 19), nor were there any differences between the groups at post-test (Table 20 and Table 21).

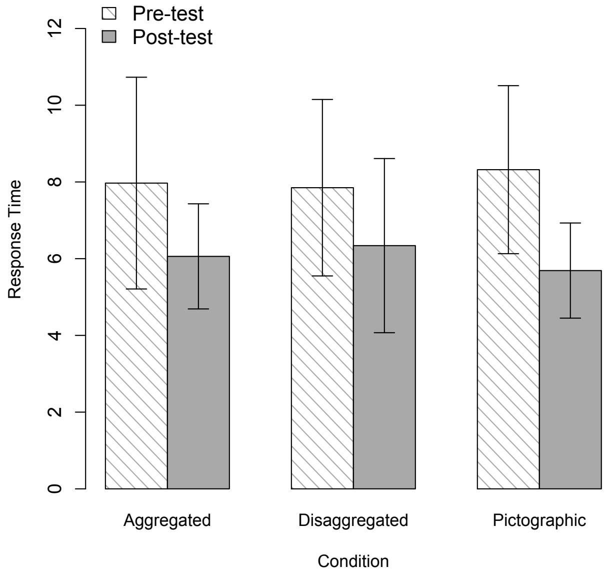


Figure 38. Response Time.

	Aggregated	Disaggregated	Pictographic	Total
Pre-test	7.97 (2.76)	7.85 (2.3)	8.31 (2.19)	8.04 (2.4)
Post-test	6.06 (1.37)	6.33 (.2.27)	5.69 (.1.24)	6.03 (1.69)

Table 18. Response Time means and standard deviations, M(SD) in seconds.

	Main effect time	Main effect condition	Interaction effect
Accuracy	F(1,62) = 5.53, p = .02*	F(2,62) = 2.66, p = .08	F(2,62)=14.75, p<.001**
Confidence	F(1,62)=51.74, p<.001**	F(2,62) = .02, p = .98	F(2,62) = 1.37, p = .26
Response time	(1,62) = 2.36, p = .13	F(2,62) = 1.51, p = .23	F(2,62) = 1.44, p = .24

Table 19. Repeated Measures Anovas.

	Pre-test	Post-test
Accuracy	F(2,62) = 1.08, p = .35	F(2,62) = 12.56, p < .001**
Confidence	F(2,62) = 2.29, p = .11	F(2,62) = .38, p = .69
Response time	F(2,62) = .22, p = .8	F(2,62) = .8, p = .45

Table 20. Anovas comparing the three groups at pre- and post-test.

	Aggregated Disaggregated	Aggregated Pictographic	Disaggregated Pictographic
Accuracy	p = .44	p < .001**	p < .001**
Confidence	p = .66	p = .39	p = .66
Response time	p = .59	p = .48	p = .21

Table 21. Pairwise comparisons at post-test with LSD adjustment.

5.15.2 Qualitative data

When asked how they made sense of the illustrated domestic electricity dataset, participants in the aggregated group reported looking at the spikes in the graph, understanding that *'the higher up the red line, the more energy was used by an object'*. However, participants in this condition had to make guesses as to how long an appliance was on for. This might have been relatively easy to guess correctly for appliances with a short duration and one high peak (e.g. the kettle), but it was substantially harder for appliances like the dishwasher, which runs for 90 minutes with two big peaks and a period of low power consumption in between. One participant found it was *'quite clear to see both the electricity power needed and the time consumed of every appliance along the timeline'*, they looked at the annotations and *'followed the spikes along the x-Axis to deduce how long the appliance has been running for'*. Yet, another participant stated: *'I was not able to understand how long each item was used'*.

Other than determining the duration of use per appliance, there was a second difficulty, which one participant described as follows: *'I tried to compare the usage over time but found this difficult because some appliances used small amounts of energy but over a longer period of time, this adds up'*. Participants in the disaggregated group could see the duration information more readily. They reported trying *'to take into consideration both axes: time and energy. So bearing in mind that even if an appliance consumes much more energy per hour, it may only be on for a short period of time'*, and they *'compared appliances which each other and memorised relationships, kettle > coffee maker; washer > microwave, for example'*. The idea that the area under the curve represented the energy consumed was clearly understood by participants: *'For the vertical axis, the higher the more it will consume per hour. Multiply it with hours. The result is the energy it has used. So, the more total space it has, more energy it has consumed'*.

Still, the disaggregated group encountered the second challenge that was identified by the aggregated group: *'it was a lot harder to compare the refrigerator, for example, as it had relatively low consumption per hour, but it was turned on for many hours'* and another explained:

> *'It was very easy to compare high-consuming appliances with low-consuming appliances, and come up with a hierarchy of consumption for high vs. low consuming appliances. However, it was harder to compare low-consuming appliances with other low-consuming appliances as the spaces on the graph were so small that it was difficult to visualise the difference between them'.*

Participants in the pictographic group only looked *'at the sizes of the boxes relative to each other'*. The duration and power information was eliminated from the data visualisation, and the cognitive task was reduced to memorising hierarchies and recalling relationships such as that the dishwasher, washing machine and fridge *'took up most energy'* and that the *'radio used least energy'*. One participant described grouping large appliances together and smaller appliances together, and then within one category compared appliances to each other to memorise their ranking. Several participants remarked that the visualisation was clear and understandable, but one participant in the group misinterpreted the boxes thinking they showed how much energy the appliances consume per hour, whereas all others understood that we displayed the typical or average duration of use of an appliance which matched *'the amount of time used in the questions'* from the energy game.

Another question asked participants what they had learnt from the information visualisation. Across conditions, several participants reported learning that certain appliances (i.e., dishwasher, washing machine, vacuum cleaner, kettle, coffee maker, microwave, and toaster) consume more than they expected. Equally, they reported surprise as to how little radio, lights and TV consume. In the aggregated condition, participants spoke mostly about peaks (*'I was surprised how much of a power spike there is for kettles'*) and they used the terms energy and power interchangeably. Only one participant in this group considers that kettle and toaster are high in power, but only used for a very short time. None of the participants mentioned the fridge.

In the disaggregated condition, the fridge generated controversial statements: several participants in the disaggregated condition listed the fridge as one of the appliances that consumed surprisingly little energy. Only one participant in this condition noted that the fridge would consume *'a lot because it is always on'* and another inferred there would always be a base rate due to the fridge. It seemed that some participants focused more on the power dimension than on the duration: *'I never thought that the fridge would consume so little electricity per hour, and I never knew that the kettle would consume that high amount of energy'*. Another participant in contrast summarised *'we should consider the power of the appliance and the length of time it works together to get the power [energy] consumption'*. Learnings were that *'more time used'* does not necessarily equal *'more energy used'*, and that *'some appliances do not have a stable consumption but it changes during the cycle of use'*.

In contrast to the disaggregated group, participants in the pictographic condition consistently stated that the fridge uses a lot. One participant realised that the fridge contributes *'greatly towards an energy bill (…) I have always been conscious as a bill payer about smaller things like charging phones, using lights etc'*. A couple of participants in the pictographic condition further mentioned that they *'learned how much energy each appliance uses relative to other appliances'*, for example, that *‘making a cup of tea was more energy consuming than listening to radio for an hour’*, which came as a surprise. The pictographic group also slightly differed from the other conditions in their assessment of the kettle and toaster – generally these were listed as appliances that were shown to consume more than participants had expected; one participant in the pictographic condition focused on the fact that *‘in total’* they use less than expected.

Participants were asked to help Jack, the persona in the scenario, with reducing his energy consumption. Most responses across the three conditions drew strongly from previous knowledge and included generic recommendations such as using appliances less and more efficiently. To use appliances less, one of the most common suggestions was to wash the dishes by hand. Others were to switch the lights off or to have fewer cups of tea and not to keep the radio or TV on in the background. To use appliances more efficiently, they recommended fully loading the dishwasher and the washing machine and boiling only the amount of water needed in the kettle.

There were a few small differences that occurred between the groups: in the aggregated group, two participants said they did not know how Jack could save energy. One requested itemised information because it would be *‘helpful for Jack to be made more aware of how much energy simple household items, e.g. microwave, use up’*. In the disaggregated condition one participant said he did not know because there was not enough information provided in the simulation, for example, the *‘kind of washing cycle’* for dishwasher and washing machine would impact energy consumption. In terms of referring to power and duration, there was one person in the aggregated group who suggested reducing the *‘using time of those with high electricity power, as these will have the significant increase on the whole electricity consumption’*. In the disaggregation group, too, some participants made the general recommendation to reduce the use of appliances with *‘high power’* or *‘high energy’*. A couple specifically named kettle, toaster, dishwasher and microwave. A subtle difference was that in the pictographic group, two participants made the same suggestion (to use appliances that need more electricity less), but they specifically named only the dishwasher and one of the two participants explicitly pointed out that the dishwasher consumes most energy (the dishwasher indeed was the biggest consumer of all appliances in the presented data).

Finally, participants responded to the question why appliances consume the energy they consume. Most participants across the conditions referred to whether an appliance is generating heat or kinetic energy as reasons why an appliance uses a lot of energy. In less technical terms, they also mentioned that the dishwasher needs power to *'clean and dry'*, the vacuum cleaner needs a lot of energy *'to suck in air'*, a radio consumes little because it only outputs sounds and lights do not have *'such an intensive task'*.

There was one theme that only came up in the disaggregated condition: A few participants referred to whether appliances need a *'boost'*, for example the kettle, the vacuum cleaner and the dishwasher do, whereas the fridge does not have a boost but it's *'on throughout the day'*. Another participant explained in a similar way that the heating element in a kettle *'causes a huge energy consumption in a short time'*, whereas the fridge is *'keeping the temperature inside instead of changing it'*. Similarly, yet another participant said: *'a light bulb just needs a little power to keep it on'*.

Again, the fridge came up as a contentious item: one participant in the aggregated group said it consumed *'very little energy per hour'* but *'across the day [it] adds up to more than most appliances'*. For one participant in the disaggregated group, the framing was exactly the other way around, i.e. the fridge *'consumed less energy although it is used 24/7'* and another said a fridge was *'designed (…) to be kept on all the time'* and *'a key feature of it was that it needed to minimise electrical consumption'*. A second participant in the disaggregated group used the same logic saying: *'Fridge need[s] to be opened [on] all day, so the consumption must be relative lower'*. In contrast, a participant in the summarised group reiterated that *'The fridge has to be turned on for 24 hours, so it consumes a lot of electricity'*.

5.16 Discussion

5.16.1 Quantitative data

The results show that participants' performance in the energy game was affected by the kind of visualisation they were exposed to and hence confirm that data comprehension depends greatly on the way data is represented (Chiang et al., 2012; Yun et al., 2010). Participants in the pictographic condition gained a more accurate understanding of how much electricity different domestic appliances were using, compared to participants who were shown time series data. This is evidence in favour of the hypothesis that area-based energy-centric visualisations are more effective for people to learn how much energy practices consume.

One might say that the pictographic representation was tailored to the task in the energy game and that therefore, it is an obvious finding that the pictographic group was most accurate. However, the purpose of Study 5.3 was to provide empirical evidence for this assumption. Study 5.2 demonstrated that simple bar charts and bubble charts were not any more effective than time series data, which seemed counter-intuitive. In Study 5.1, we found that the normalised condition was slightly better than the aggregated time series, however the effect was very small (Figure 25). In Study 5.3, the difference between the time series graphs and the pictographic condition is unambiguous (Figure 36). The pictographic group had approximately 10 correct comparisons more than the other groups in the post-test (this is almost an 18% difference between the aggregated and the pictographic group).

The other part of the hypothesis, namely that participants seeing the disaggregated time series graph would be more accurate than participants seeing the aggregated graph, must be rejected again. Neither Study 5.1 nor Study 5.3 found any evidence that disaggregating the data and displaying it as multiple colour-coded line graphs makes it any easier for people to learn how much energy appliances consume than an aggregated curve showing total consumption. This demonstrates that disaggregating energy data alone does not help to identify where energy is spent, it only becomes easier when represented in an energy-centric manner.

We did not find meaningful differences between the conditions for the variables response confidence and response time. We ran these additional measures because we looked to previous publications and these have used accuracy, time and confidence. We expected that with higher accuracy, participants would also be more confident and give quicker responses if they felt they knew the answer, as opposed to hesitating to answer if they were uncertain. All three groups increased their confidence score and decreased their response time from pre- to post-test. Seeing that the aggregated and disaggregated condition did not improve their accuracy score (participants in the aggregated condition even decreased their descriptive accuracy score from pre- to post-test, even if not significantly), it seems that becoming faster and more confident could merely be an effect of repeating the energy test (however, we discuss other possible reasons in 5.16.2). The finding that participants in the time series conditions became more confident (albeit not wiser) is further evidence that it is difficult to correctly extract information about cumulative energy consumption from time series graphs.

5.16.2 Qualitative data

The findings from the qualitative data are in line with previous studies (Chisik, 2011) and with the data from Study 5.1 and Study 5.2. In the simulation, participants learned that some of their assumptions were biased by flawed heuristics (Attari et al., 2010). As expected, challenges occurred in the disaggregated condition in working out power over time, and in the aggregated condition even more so as participants focused on the amount of power but could only make guesses about duration. The comparisons were particularly difficult for appliances that did not differ greatly in the energy they consume, whereas participants found it feasible to work out the hierarchy between appliances with sufficient differences.

Interestingly, several statements in the qualitative data show that participants thought they learned from the line graphs. It seems they overestimated their ability to work out the duration in the aggregated condition and the respective area under both types of line graphs. This might explain the increase in confidence and decrease in response time, because participants subjectively felt they learned from the visualisation, only, objectively they did not (as evidenced by the accuracy data). In the pictographic condition on the other hand, confidence went up because participants were certain that they learned and indeed they did. Most of the responses to the question how Jack could save energy were generic and similar between the conditions, one subtle difference was that only in the pictographic condition a participant verbalised the explicit link between the dishwasher being the top consumer and his recommendation for Jack to use it less.

Particularly striking are the responses relating to the fridge. Participants in the pictographic condition identified the fridge as a big consumer, explaining this by saying it is on 24/7. The explanation is obviously based on prior knowledge (because the time information was not provided in the visualisation), but it seems that the size of the fridge icon was memorable. The disaggregated time series condition should come to the same conclusion about the fridge, as it is clearly visible in the graph that it is always on, adding up to a significant amount of energy over time. Yet, they failed to do so, saying the fridge consumes surprisingly little even though it is constantly on.

We were expecting to find the time series graphs to trigger deeper reflection than the pictographic visualisation, but we cannot necessarily confirm that this was the case. However, it seems that the additional questions we asked in Study 5.3 are a good approach to probe people's reasoning beyond the data and to explore their energy literacy. While the quantitative data only provides an objective score, the qualitative data can help explain which explanations participants constructed to memorise the data, and it further explains how they arrived at erroneous conclusions (e.g. that the fridge consumes little energy). Participants giving only generic recommendations on how Jack could save energy may be interpreted as low energy literacy. On the other hand, if participants fully comprehended the data, they were able to give specific advice, reflecting higher levels of data literacy and energy literacy.

5.16.3 Limitations

The limitations discussed in Study 5.1 (5.6.3) still hold for Study 5.3 (i.e. lack of generalisability due to a mostly female student sample and limited ecological validity).

In addition, Study 5.3 put into question whether response confidence and response time were meaningful measures. In Study 5.1, we interpreted the increase in confidence as an indicator that the normalised group was more effective. In the last experiment, it turned out that all groups became faster and more confident, independent of whether they learned or not. An improvement would be to record the time participants spend looking at the data visualisation instead, to measure how long it takes them to understand and memorise the information.

5.17 Conclusion

The third experiment provides evidence in favour of area-based energy-centric visualisations that deemphasise time. It showed that pictographic blocks representing energy consumed by individual appliances lead participants to answer significantly more comparisons correctly in the energy game. This suggests that appliance-level disaggregation in an easy to comprehend representation is significantly better suited for people to learn how much energy everyday practices consume.

5.18 Introduction to Study 6

Chapter 4 identified challenges with current smart energy visualisations through interview studies with householders in the field. The lab experiments in Study 5 systematically tested six types of visualisations (aggregated line graphs, disaggregated line graphs, normalised line graphs, bar charts, bubble charts, and pictographic charts) and found the best results for area-based energy-centric representations using pictographs of household appliances. The purpose of the this study is to take the findings from the controlled lab setting back into a field setting and to test how users in the real-world context respond to the pictographic visualisation.

Study 6 is a field trial that tests a visualisation which aims to integrate the previous results by providing energy data feedback that is appliance-centric and interactive. The basis for the visualisation are pictographs that are scaled in size to mirror the energy consumed by individual appliances. Visualising information through the size of an area representation has been attempted in FigureEnergy in Study 4 (and before in Costanza, Ramchurn, & Jennings, 2012), and has proven effective in Study 5.1 and 5.3 and before in other works (Borghouts et al., 2015; Schwartz, Denef, et al., 2013).

While energy-centric visualisations are best suited to convey how much energy appliances or practices are consuming, they miss out on any time-related information. Study 5 found some evidence that it was time-centric graphs that triggered deeper reflection. Time-related information is immediately relevant to evaluate one's energy consumption, as both power and duration of use determine energy consumption. For example, time-related information like idle time of appliances should be given to householders according to a recent review of effective energy feedback interventions (Murugesan et al., 2015). In the real-world, time-specific information will be relevant in many use cases, for example in demand-side management, where customers might wish to shift their energy consumption to a time when they pay a lower rate for their energy.

Ideally, a visualisation would provide both energy-centric information and time-centric details, like FigureEnergy does (Costanza, Ramchurn, & Jennings, 2012). A problem with FigureEnergy in Study 4 was that the Consumption Graph and the Consumption Overview were in different view tabs, and participants spent most their time with the Consumption Graph and did not use the Consumption Overview. The challenge is how to fit both types of information into one visualisation. Figure 39 shows a visualisation by energy provider Fresh Energy, showing pricing information of energy use directly attached to pictographs of appliances. The pictographs in this example are not scaled in size to represent energy consumption, but the attached labels create the opportunity to add further numeric information to the visual representation.

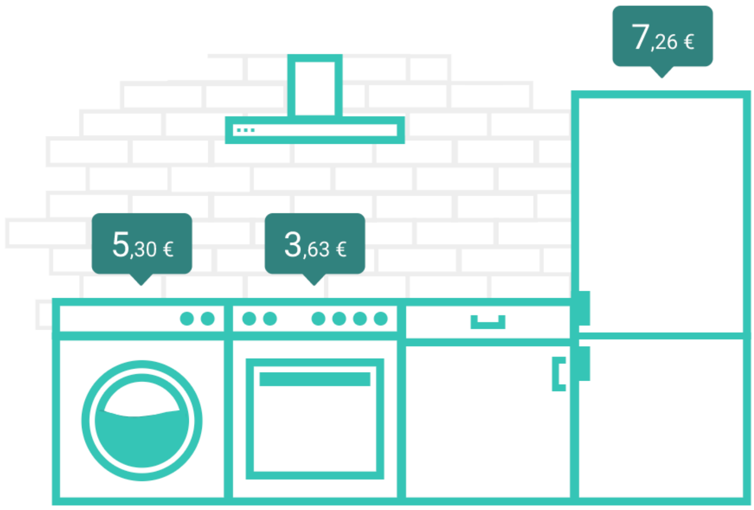


Figure 39. Appliance-level visualisation by provider Fresh Energy.

The feedback visualisation deployed in Study 6 combines energy-centric and some time-centric information in a prototype that displays energy-centric data in form of area-based pictographs, and adds some time-centric information through an interactive feature. The pictograph visualisation is equivalent to the pictographic condition tested in Study 5.3. Each pictograph showed the energy consumed by an appliance through its size. It featured interactivity that allowed users to click on each pictograph to reveal more information. This would open a pop-up window displaying details on how often and how long the appliance was used for. Interaction is considered a crucial feature in Info Vis and it has been found to be an important feature of energy feedback systems by supporting active exploration increasing engagement and reflection (Study 4; Costanza, Ramchurn, & Jennings, 2012; Munzner, 2014; Sedlmair, Meyer, & Munzner, 2012). The aim of the deployment is to observe users' reflection on the feedback, whether they connect it to their lives, and thereby to evaluate the visualisation.

In this field deployment, appliance-level data was collected in eight households and visualised through a web-page. Participants took part in a contextual inquiry (Holtzblatt & Jones, 1993) using the think-out-loud method (Lewis & Rieman, 1993), i.e. householders talked through the information displayed on a web-page. All interviews were audio recorded and transcribed in the transcription software f5. The transcripts were imported into the qualitative data analysis software Nvivo and analysed thematically (Aronson, 1995; Clarke & Braun, 2014; Seidman, 2013).

5.19 Method

5.19.1 Sample

For this field study, eight employees (one female) of EDF Energy UK were recruited. The mean average age of the sample was 41.7 years (SD = 7.3 years). All participants lived in south-east England, some were renting and others owned their houses which ranged in size from 3-4 bedrooms. Participants reported they spend between £40 and £115 per month on their electricity bill. A couple of them were unsure if this figure includes gas and remarked that they found their bills were unclear. Participants received their bills in different formats varying between conventional paper bills and digital bills received via email, web account or mobile app. Three participants did not report any smart technologies in their home (other than the smart plugs installed for this study). Five had smart technologies such as smart thermostats, smart lights, Amazon Echos, a Nest camera, and a robot cleaner.

5.19.2 Material

All households were equipped with a current clamp meter to record overall electricity consumption in the house. Furthermore, smart plugs tracked the consumption of individual appliances. The number of smart plugs per house ranged from two to six smart plugs. Table 22 shows an overview over the appliances that were equipped with smart plugs in the eight households (participants P1-P8). All sensors were installed by collaborators at EDF and participants could choose the appliances they were interested in tracking. The chosen appliances remained static (they were not moved around the house) and no other appliances would be plugged into the assigned smart plug during the study.

	P1	P2	P3	P4	P5	P6	P7	P8
Washing machine		X		X	X		X	X
Kettle	X	X			X		X	X
TV/Entertainment				X		X		X
Fridge	X		X	X		X	X	X
Dishwasher	X		X			X		X
Microwave						X		X
Tumble dryer	X	X						
Washer dryer			X			X		
Toaster							X	
Hair dryer							X	

Table 22. Smart plugs installed on participants' appliances.

Figure 40 shows the design of the web interface. The default view shows the information for the 'Last 7 days' (top right corner). At the top of the page, the user can select dates to look at only one day, several weeks, or any specific time episode of their choice.

In the top half of the page, each appliance that is equipped with a smart plug is visualised as a two-dimensional pictograph, the area corresponding proportionally to the amount of energy consumed by the appliance. The user can interact with the visualisation by hovering over the icon, thus revealing appliance label, how often the appliance was started, the total time of use, and the total energy consumed. Other 'always on' appliances are shown at the bottom of the events visualisation.

In the lower half of the page, a bar chart in the left corner shows the same information as the box view above, i.e., the energy consumed per appliance. In the right corner, an appliance-specific comparison tells the user if appliances have consumed more energy or less energy in the current week compared to the previous week.

Figure 40. The web interface showing participants' energy consumption.

5.19.3 Procedure

Before the contextual inquiry, participants answered questions assessing their energy literacy:

1. *On a scale from 1-5, how much do you feel you know about your energy use at home?*
2. *Electrical power is commonly measured in units of?*
3. *Electrical energy is commonly measured in units of?*
4. *A lightbulb (CFL) uses 13 W. If it is run for 2 hours, how much energy does it use?*
5. *Please pick five appliances in your household and rank them in terms of how much electricity you think they consume and explain why.*

In the contextual inquiry, participants were asked to log in to the website, and to think aloud about the feedback they saw. When they stopped thinking out loud, they would be prompted with questions (e.g. What do you learn about your energy use? Can you please go through a few days and compare? If you wanted to reduce your consumption, what would you do?). The interviews were conducted via skype or phone and lasted from 15 to 35 minutes with an average duration of 21.5 minutes (SD=7). Participants would share their screen via Skype or the researcher would be logged in to the account at the same time to allow them to see what the participants was seeing.

5.20 Results

5.20.1 Energy Literacy

Participants were asked to rate themselves on a scale from 1 to 5 with regards to how much they felt they knew about their energy use at home (1 = very little, 5 = a lot). On average, they rated themselves M = 3.83 (SD = .75; min = 3; max = 5). They were also asked to answer three items assessing energy literacy. All but one participant answered all three items correctly, one participant answered two items correctly.

Furthermore, the questionnaire asked participants to pick five appliances in their household, to rank them in terms of how much electricity they consume, and to explain why. This was both to get another estimate of participants' literacy, as well as to see if the recorded energy data differed from participants' expectations. All lists were assessed as reasonable and were occasionally used in the interview to prompt reflection (e.g. P1: tumble dryer/washing machine/ fridge/kettle/coffee machine).

5.20.2 Contextual Inquiry

The data from the contextual inquiry was analysed thematically and is presented below structured by the following five main themes that emerged. First, we report the evaluation of the chosen data visualisation (the pictographic blocks representing energy use). Second, we report participants' sense making, i.e. we found they reflected on the power usage of appliances, their duration of use and their frequency of use to make sense of the total consumption and to compare between appliances. Third, we report data describing participants' reasoning about changes in consumption over time. Fourth, a theme emerged around information that was currently invisible, i.e. participants would like to learn more about the individual use of appliances to do with their habits and the settings they use. Fifth, building on that, participants spoke about opportunities for behaviour change and other efficiency measures to optimise energy consumption.

5.20.2.1 Prototype evaluation

The focus of the prototype evaluation addresses the area-based pictographs. Other usability issues of the prototype are reported at the end of this section.

Five participants recognised instantly that the two-dimensional boxes represent the energy consumed by the depicted device:

> *'Immediately you recognise there was an information shared through the size of the icon (…) A bit like in a presentation with bubbles, but with the icons it's obvious, because you don't have to refer to a key [legend] to understand, you just see the appliance and you understand the size is how much it's used' (P2).*

The boxes allowed participants to quickly and effortlessly compare consumption: P7 and P8 spontaneously listed their appliances in ranking order from high to low; P6 identified the highest consumer correctly when prompted by the interviewer.

P2 commented on the visualisation as *'neat'* and was *'impressed by the simplicity (…) of the way the information is displayed'*. P2 and P3 thought it was *'excellent'* (P2) and it *'stood out'* (P3) that the pictographs' size changed when they changed the dates they were looking at. P2 elaborated that the relative consumption of the washing machine and the tumble dryer changed when he changed the dates. P7 when changing dates, could confirm easily that *'The overall usage stays the same, in terms of the order'.*

Three participants did not instantly see that the pictographs' size represented energy. For example, P5 approached the visualisation very analytically, thinking out loud if it could be a bar chart without scales, he wondered if the water levels depicted in the kettle and washing machine have a meaning, before concluding that it is *'one of these square diagrams where the size of the box represents the amount they are… the total consumption'*. He adds: *'Usually I see those sort of being one rectangle where the various blocks within the rectangle fill up the whole rectangle rather than (…) not filling up the whole area.'* Likewise, for P4 the purpose of the boxes became obvious only after he changed the dates and the boxes changed accordingly: *'Oh, actually the size shows the consumption! That's good!'*. In contrast, P1 said he was used to working with data in his job, he looked straight at the numeric information in the pop-up windows and did not notice the relevance of the icons' size.

To make the meaning of the pictographs' size more obvious and to facilitate comparisons between appliances, participants voiced the request to have the appliances sorted in order ranking from high to low or from low to high (P1, P7), whereas at the time they were presented in random order. This ranking was provided by the bar charts at the bottom of the page, put participants did not pay much attention to them. It is assumed that the bar chart was redundant as the same information was conveyed through the main visualisation above.

One problem with the prototype was that the 'time started' count was confusing in the case of the fridge which was perceived as *'always on'*, even though the fridge's cooling comes on sporadically. P1 and P7 judged the number of 'times started' as incorrect (too high) for the dishwasher and the washing machine respectively.

5.20.2.2 Appliances' power, duration of use and frequency of use

Participants used the pictographs and the pop-up information intuitively to compare between the appliances equipped with a smart plug and to identify which appliances consumed the most electricity.

P1 found that his tumble dryer was *'a killer'*. Further, he realised that his *'fridge is a bit of a killer'*, too, seeing that it took only a little less than the tumble dryer. He realised that the fridge, although less power intensive than the tumble dryer, amounted to the same energy use as the tumble dryer because the fridge is always on.

P4 was *'surprised that the TV takes a lot more than the washing machine and the fridge'* and concluded that this is due to habits: *'We watched too much TV then I guess'*. Indeed, for the week in question, the TV had been on for 23 hours and 51minutes, and for the whole 28 days of February, it had been on for 8 days, 5 hours, and 40 minutes. He elaborated that his flat mate *'has a bad habit of sometimes leaving the TV on when sleeping.'*

P1 was *'surprised'* that the kettle did not consume more total energy because he *'always thought that the kettle takes a lot'* of power. Vice versa, other participants found that their kettle and toaster took a bigger share than expected. P5 found his kettle was on for a total of 2 hours and 30 minutes: *'That's a lot! For a kettle that normally doesn't take very long to run.'* P8, too, was *'surprised that over the course of a month the kettle is on for over 2 hours'*, given that *'if you're boiling the kettle for one drink it probably takes 2 minutes'*. P7 found out that *'Surprisingly, it's [the toaster has] taken (…) 10 minutes, I thought that would be less than that'.*

Engaging with the numeric data in the pop-up, P2 said that it helps him *'to figure out if it's accurate'*. It provided insights into the usage of appliances, for example, the washing machine has been started 43 times, whereas the tumble dryer was started only 15 times – despite its lower frequency it's runs for longer periods of time and therefore came close to the consumption of the washing machine.

Similarly, P1 and P3 reasoned:

> *'tumble dryer, almost 50kwh, and it started 32 times. So that must take a lot more energy than the fridge, which started 839 times. I only do it by comparison' (P1).*
>
> *'So I can see I have used… more than double the amount of energy on the washer dryer than I have on the fridge. In a fraction of the number of times. It makes sense. I know the washer dryer uses more energy than a fridge' (P3).*

P5 found it *'interesting that the kettle, even though the washing machine is a bigger appliance and probably runs for longer, that the kettle actually needs the same amount'* when added up over time. Doing the maths for the washing machine, which ran seven times for a total of 7 hours and 27 minutes, he inferred that they run approximately one-hour cycles in general. Elaborating on the use of the kettle, P5 said *'I guess it would be interesting to understand the average amount per use, in kWh (…) that is the final piece that basically helps you reconcile the two'.*

Participants mentioned two user requirements. First, they voiced a need for references as they did not know how much appliances should be using, i.e. they couldn't get a feeling whether their use in kilowatt-hours was low, normal, or high. P3 thought she was *'probably fairly typical in the fact that I don't have an easy way of saying 1 kwh is that good or bad?'.*

Second and related to that, participants would like to see cost in Pounds in addition to consumption in kilowatt-hours. P1 said: *'I understand kilowatt-hours. The thing is without monetary value it just doesn't tell me a lot'*. P2, too, would have found that *'useful'* and added that he always feels it would be good to explain kilowatt-hours to people in an abstract way:

> *'what one kilowatt-hour would look like should you have to kind of make a physical effort to deliver it (…) kind of muscular effort, to produce 1kWh (…) how many km of jogging would you have to do to cover the equivalent of 1kWh. Just to make people realise that 1kWh is a lot of energy.'*

5.20.2.3 Changes in consumption over time

Most participants explicitly linked the consumption information to their awareness and knowledge about their energy use in the home. P3 reflected on the high consumption of the washer dryer saying they had a *'big washing week'*, and P2 explained that in a certain week *'we used the tumble dryer a lot'*.

Looking at the changes in consumption between 'this week compared to last week', P6 attributed the 35% increase in microwave usage to having friends over and preparing lunch and dinners for them. Similarly, P7 explained the decrease in the use of his kettle saying that in the previous week his *'wife had a cold, she would have done hot water bottles'*. For the change in the fridge's consumption he wondered if *'maybe what we have in the fridge has changed a little bit'*.

P1 thought 'this week compared to last week' made him think *'What happened this week?'* and he thought *'you could map it against your family calendar and you can see what events took place and what makes you consume more energy – "Oh that week we did massive shopping, wow, the fridge's consumption has gone up"'.*

Sometimes participants did not have an explanation for the change in consumption between the previous and the current week. P5, who had used 59% more electricity for the washing machine in the last seven days compared to the week before, wondered *'whether last week was a typical week or a quiet week'*. P3 would have liked to know why it was that *'the fridge has used less one week compared to the next'.*

P6 concluded:

> *'The other problem I have is that the use of electricity for the microwave has increased by 35% but it's quite small because the microwave represents only, yes, 5% of the total electricity use. So maybe what is missing here, you have the relative impact, 35% for the microwave, what is the absolute impact? On the electricity consumption, or the electricity bill? Basically, to be able to identify what is really important and what is less important.'*

5.20.2.4 Requests for more information

Participants requested additional information that they could not see from the data, but would like to learn about. These were seeing more devices than were currently equipped with smart plugs, more details about how habits in the home affected their energy consumption (such as running devices on different settings or programmes), and actionable advice on how to re-evaluate their energy use.

Firstly, when asked if there were other devices in their home they would like to see that had not been fitted with smart plugs, they listed electric hobs, ovens, microwaves, and fridges as assumed big appliances. They were also interested in *'small things'*, such as media and entertainment kits, the broadband router, the cable box, Amazon Echos, charging phones, lighting, coffee machines, TV, or specific appliances like a fish tank. Most of these *'small'* things fell under the always-on category, which P3 referred to as *'noise'*, and P6 as *'difficult to track'*. P2, too, thought they are *'a bit of a challenge'* and he wished for a better understanding of how much they consume. At the same time, several participants remarked that there was little control over these appliances, that they had to be on stand-by for convenience. P3 pointed out she would not consider replacing the fish tank.

Secondly, participants wanted to know more about the impact of their habits on consumption. For example, how much energy they were using in each room (P4) or how the individual use of the kettle would change its energy consumption. P5 wanted to know how full the kettle was each time, how long it was boiling for and how long the intervals between two boils were *('I know sometimes we would boil the kettle and then not make something immediately and then boil it again'*). Similarly, P5 would have liked to see what the maximum power of all lights in the house was and how often they reached the maximum. He wanted to see if they switch lights off when not needed or if they often keep all lights on.

Thirdly, they were looking for actionable advice. P1 wondered if the website *'could give tips on how to best use the dishwasher and the tumble dryer'*. Indeed, P5 puzzled over the settings of his washing machine and dishwasher, wondering which programmes are efficient:

> *'quite often we put [the washing machine] on a fast cycle because it's quick and I have no idea if putting it on a three-hour cycle actually uses less energy in the end (…) Similarly with the dishwasher (…) eco-mode is a three-hour cycle. Whereas I could put it on a fast cycle and it might actually… the eco-*

element might just be the water consumption rather than the energy consumption.'

P8 thought of eco-feedback as a two-step challenge: *'Explaining where the energy goes is one thing, but then helping them decide how they can reduce it is another.'* He considered tips such as *'don't overfill the kettle'*, *'make sure the washing machine is full'*, *'Make sure you're washing 30 degrees instead of 40'* as solutions for the second step. In the following section, participants' thoughts on behaviour change and efficiency measures are presented in more detail.

5.20.2.5 Changing habits, settings and other efficiency measures

The interview data suggests that participants saw a range of opportunities to increase energy efficiency. In terms of habits, P4 identified a behaviour that he would like to change based on the data feedback: *'clearly the habits… less TV, telling my flatmate not to keep the TV on while sleeping (laughing) surely.'* P8 spoke about the efficiency-measures that they already had in place, such as making use of the dishwasher's eco-programme: *'it runs over night so we don't really care that it takes four hours'.* P6 and P7, too, considered revisiting the operational settings of their appliances:

> *'Maybe we can change the program. Using a lower consumption programme for the dishwasher' (P6).*
>
> *'[the fridge is] already on a low setting but maybe we could adjust that a little bit and see if we can make the comparison in terms of last week versus the next week better' (P7).*

P8 pointed out that a solution other than changing behaviours and settings would be to purchase more efficient appliances in the first place. P2 emphasised that they taught the children to switch the lights off and not to boil more water than needed, but he went further in implementing smart solution where possible. For example, he had movement detectors in the corridor to control the lights automatically.

Participants also spoke about limitations for behaviour change and energy savings. P8 suggested that making pasta in the microwave instead of the pan would be more energy efficient, but *'Nobody's gonna do that'*. In P2's words:

> *'lifestyle is not really negotiable (…) when you work you (…) deserve a minimum kind of standard. Which means I am able not to worry all the time about anything being switched off (…) Because your life would become a nightmare. And you have to make sure everybody is comfortable living in the house.'*

Similarly, P7 said they could not stop using the hairdryer in winter and his wife would say *'if I have a device (…) and it's more efficient, then that's probably the best way (…) Whereas if you're asking me to undercook the meal then it's not gonna happen.'*

5.21 Discussion

5.21.1 Main findings

The main finding from this study is that the appliance-specific data collected by the smart plugs allowed actionable insights about how energy is used in the home. Data visualisation literature suggests that pictographs can have benefits for easily engaging people (Haroz et al., 2015) and indeed, the pictographic blocks enabled quick comparisons between devices and the identification of the appliances using most energy. Participants gained practical insights (such as finding out that the TV was running a lot and hence consumes a big share of the total energy in the home), which enabled them to re-evaluate practices in the home.

The findings are consistent with previous research. For example, home energy use is a complex environment with multiple actors (Busby & Chung, 2003; Osman, 2011) and social conflicts may arise when one person consumes more than their fair share (Leygue, Ferguson, Skatova, & Spence, 2014). When it comes to optimising energy use in the house, people weigh up the benefits of changing their behaviour to save energy and the inconvenience that would cause. The idea of deserving a certain standard reflects what has been found in previous work, namely that many aspects of personal lifestyle are non-negotiable, and that being comfortable in the home is vital (Hargreaves et al., 2010; Shove, 2003). The findings also resemble what was found in Study 4 with regards to waste. While waste is very subjective, people tend to be innately averse to the idea of squandering energy.

5.21.2 Reflection

The pictographs could be clicked on and would open a pop-up window with information on frequency of use, duration of use, and energy consumed. Participants naturally interacted with this feature and reflected on how often they use appliances and which programmes they run. The information provided in the pop-ups helped with three kinds of reflection about energy consumption in the household:

1. Identifying appliances that consume the most electricity
2. Reflecting on the relative impacts of power, duration of use and frequency of use
3. Identifying behaviour-centric and appliance-centric potential for energy savings

The interview data revealed a common theme, namely interviewees thinking about the interplay between power, duration per use and frequency of use. Previous work has focused on energy as the product of power over time, which is the physical definition of energy. Total time of use can be split into duration per one use and frequency of use. The first (duration) is mostly determined by appliance-specific aspects, e.g. the washing machine's chosen programme, or the amount of water in a kettle. The second (frequency of use) is determined by habits, i.e. how frequently someone washes laundry or drinks tea. Only at this level of distinction between time per use and total time does the data map to everyday practices. The information of how often and how long appliances were used and how much energy they consumed helped correct flawed assumptions that participants had about the energy usage of appliances (Attari et al., 2010).

These findings provide evidence with regards to how users reflect about home energy use and ways to change it, given the available appliance-specific data. Participants considered how power, duration of use, and frequency of use determined energy consumption. First, power usage could be manipulated by retrofitting appliances with more efficient ones, by optimising the settings of an appliance (e.g. set the fridge to a cooler temperature), or by choosing a different programme (e.g. washing laundry with the eco-programme). The duration of use can be influenced both by choosing a more efficient programme (e.g. eco-wash), and by changing practices (e.g. switching the TV off when going to bed). The frequency of use can only be influenced by practices (e.g. waiting to make tea instead of forgetting the kettle and re-boiling). Whilst replacing appliances (Attari et al., 2010) might not be an available option to every household, simple changes in the operation of devices can be made relatively easily.

The current prototype provides the total kilowatt-hours consumed per appliance, but not the average power or energy consumed per individual use. This piece of information was missing in the prototype to complete the picture and to support users' reflection. In addition, more information could be added to make the data more meaningful. For instance, kilowatt-hours are an intangible unit and participants in this study asked to see the equivalent in Pound sterling, which is in line with the findings from Study 3 and other work that indicates that units other than kilowatt-hours (Pounds, or CO^2) are more beneficial (Spence, Leygue, Bedwell, & O'Malley, 2014).

5.21.3 Limitations

All but one participant answered all technical energy literacy questions correctly. Yet, they only rated themselves M = 3.83 on average on a 5-point scale on how much they felt they knew about their energy use. This has two implications. First, the self-assessment of energy literacy (Yun et al., 2011) does not seem to be a suitable measure, if employees of an energy company who can be considered energy experts, rate themselves at M = 3.83 (for comparison, householders in Study 2 rated themselves M = 3.4). Second, the nature of the highly energy-literate sample limits the generalisability of the findings. The study would need to be replicated with a sample of non-experts. It is worth noting that even though the sample consisted of what can be considered energy experts, they were comparable to the average end-user in saying that they typically do not engage with their bills and seven out of the eight participants said they did not feel that the statistics expressed in kilowatt-hours were meaningful to them. Nonetheless, they were tech-savvy early adopters who already owned several smart home technologies which sets them apart from the general population.

The other limitation is that smart plugs were installed selectively at appliances that were accessible and chosen by participants. For a more holistic assessment of disaggregated data feedback it would be desirable to provide householders with feedback for all appliances in their home. Due to the website being a prototype, a few usability issues emerged with regards to the appliance usage data. The web-page reported the fridges' cooling cycles as 'times started', which was confusing for participants, because this contradicts the mental model of a fridge being always on. It is likely that the count for washing machines and dishwashers was inaccurate – these two appliances typically have two high peaks and low power consumption in between. Participants perceived the 'times started' count as too high and it is possible that the smart plug counted two starts that were part of one programme.

Finally, this was the first deployment of a prototype. The interactivity of the feedback was very basic and the feedback did not involve information about peak times of energy use. These are aspects that need further investigation in future research. With services such as time-of-use tariffs incentivising load shifting, time-centric information is highly relevant for householders and needs to be an integral component in smart energy feedback.

5.22 Conclusions

The visualisation tested in Study 6 combined the advantages of the aspects identified as effective for disaggregated feedback in the previous studies. Participants were given an area-based energy-centric visualisation using pictographs, which they could interact with to reveal additional information about frequency and duration of use. In comparison to the previous studies, this visualisation evoked substantial reflection in participants and provided actionable insights on how to re-evaluate practices in the home. The study provides evidence that disaggregated data is useful and necessary for householders to learn how they are consuming energy.

5.23 Conclusion from the Visualisation Studies

Study 5 and 6 addressed RQ3: How does the design of the data visualisation affect how people make sense of domestic energy data? Study 5 systematically tested visualisation configurations that were identified as interesting to investigate in Study 3 and Study 4. In Study 5, a set of three experiments evaluated the advantages and shortcomings of different energy data visualisations and identified area-based energy-centric visualisations as the most suitable ones for householders to learn how much energy everyday practices consume. This visualisation was then tested in Study 6, a final field study, to confirm if the findings of the experiments would be useful for householders analysing their actual energy consumption data in the real world. The field study validated that the energy-centric visualisation, in combination with interaction (even if it was very basic interaction) and some time-related data, allowed householders to learn how they are using energy and to re-evaluate practices in the home to optimise energy consumption.

Chapter 6 Discussion

6.1 Summary of Findings

The main research question in this thesis was: Do householders understand smart electricity feedback? This main research question was addressed through a sub-set of three questions and a mixed methods approach (Cairns & Cox, 2008). First, the three sub-questions, the methodology to address them, and their findings are summarised.

RQ1: What is energy literacy?

The first research question sought to define energy literacy in the context of home energy feedback. The method chosen to discuss and redefine energy literacy was a set of three focus groups conducted with energy experts from academia, energy experts from industry, and energy-customers. Based on this study, we define energy literacy as knowledge about energy (not including attitudes or behaviours). In the context of home energy use, we define actionable energy literacy as practical knowledge about how much energy practices in the home consume.

RQ2: How do householders interact with smart electricity feedback?

The second research question explored how users interact with and make sense of smart energy feedback. This included interviews with householders who had Smart Meter In-Home Displays, householders who were using a commercial web-based smart feedback tool, and householders who were using a web-based research prototype. The results indicate that people learn little from smart energy feedback systems that are currently available on the market. They often do not see the link between the data and how they are using energy through practices in the home. Many current generation IHDs or web-based systems focus on giving feedback on instantaneous use, and a summary of historical use. However, people said that these features were not very useful because they do not provide information about the energy consumed for specific activities.

One factor that helped to make time series data more meaningful was active interaction and manipulation of the data (i.e. keeping a digital diary within the feedback software FigureEnergy in Study 4, or clicking on the information visualisation to retrieve additional information in the EDF field trial in Study 6), which enabled householders to reflect more deeply and to re-evaluate their practices. Only participants in Studies 4 and 6, which both had basic interactive elements and focused on appliance- or event-centric data, triggered re-evaluations of practices in participants. The two field studies without interactive elements or itemised feedback (Studies 2 and 3) did not find examples of participants reflecting deeply on their habits, the setting of their appliances, or whether they had any wasteful behaviours that they wanted to address.

The findings indicate that feedback on the total consumption of a household was not useful, and that householders require disaggregated data on the appliance-level to understand how much energy practices in the home consume. Further, based on these findings, we predict that feedback systems with more sophisticated interaction would further stimulate reflection and data comprehension. Study 6 provided statistics on duration and frequency of use, and participants in this study voiced requests to interact more deeply and retrieve more detailed information. Only then they said could they fully explain the data.

RQ3: How does the design of the data visualisation affect how people make sense of domestic energy data?

The third research question focused on testing whether visualisations that show disaggregated appliance-level data, can increase people's understanding of how much energy everyday activities in the home consume. This was first tested in a set of three lab experiments, which found that disaggregated, activity-centric feedback was significantly better than aggregated feedback for helping people to learn how much energy everyday activities consume. However, it is important to mind that disaggregation is necessary, but not sufficient, as the evidence from Study 5 shows: there was no difference between aggregated and disaggregated time series line graphs. Time-centric energy data visualisations were not as useful as an energy-centric visualisation, which summarises the energy consumed by a practice in an area-based visualisation. A visualisation with energy-centric pictographs, indicating energy use by the size of the pictograph, was then deployed in a field trial. The web-based visualisation in this trial added an interactive element (which was identified as useful in Study 4) to provide participants with information on the frequency of use and the duration of use of appliances. The study found that this visualisation was useful for people and they learned how much energy everyday activities in the home consume.

Main RQ: Do householders understand smart electricity feedback?

The answer to the main research question – Do householders understand smart electricity feedback – is that householders' understanding of current generation smart energy feedback systems is limited. They often cannot explain the data and they cannot link it to everyday practices in the home, which means they do not learn which practices in the home contribute most to their energy consumption. Neither do they identify potentially profligate practices which could be re-evaluated and changed. For householders to gain these insights, energy feedback needs to provide appliance-centric information.

6.2 Contribution

First, this thesis has provided an in-depth analysis of how householders interact with and make sense of smart energy feedback. It has identified that suitable design of the feedback is essential for data comprehension. This work links research from the domain of data and graph comprehension with personal informatics theory. Theoretical models from the domain of personal informatics (Epstein et al., 2015; Li et al., 2010) emphasise the central role of reflection when people engage with data. Reflection has been found to be crucial for people to integrate new information into their existing knowledge structures and to be able to identify opportunities for behaviour change (Ploderer et al., 2014). If barriers arise in the reflection stage, people cannot move on to the later stages and will not change their behaviour (Li et al., 2010).

This thesis has added to the understanding of the role of reflection and it has demonstrated that research needs to focus on the users' cognitive processes when engaging with data. Prior studies have focused heavily on the outcome of energy feedback (i.e. they have measured savings), and have disregarded Epstein et al.'s (2015) insight that data is not always collected with the single purpose of behavioural change. A strong motivator for people to engage with data is curiosity. However, their curiosity will be stifled if they do not find sufficiently relevant and interesting information in the data. For energy feedback, this thesis has demonstrated that relevant and interesting information must map to the social context, i.e. it must relate directly to activities in the home. Based on the above-mentioned models, a lack thereof will pose a barrier at the reflection stage and prevent users from progressing into the later stages.

Second, this thesis provides evidence that disaggregated appliance-level information is more useful to householders than aggregate feedback. This contributes to the body of research indicating that householders need disaggregated data (Murugesan et al., 2015; Neustaedter et al., 2013). A simple technological solution to obtain appliance-level data has not been found yet (Batra et al., 2014; Zeifman & Roth, 2011), but the evidence in this thesis provides a proof of concept that disaggregation is useful. The grocery bill metaphor from (Kempton & Layne, 1994) is still relevant today, because even the real-time feedback offered by Smart Meters does not provide an itemised overview of the cost of running each appliance in the home. Moreover, Kempton and Layne's observation still holds true that conclusions that householders can draw are limited both by how they receive information and by their (possibly limited) analytical capabilities.

Third, in providing disaggregated data, it is central to consider human cognition, mental models, and visual processes involved in making sense of visualised data feedback (Cheng & Barone, 2017; Cleveland & McGill, 1984; Pinker, 1990). This thesis has collected data that illuminates how people make sense of energy data visualisations, and why seemingly suitable graphs that are in keeping with graph theory are still difficult to comprehend for householders because they conflict with the social reality of energy use and its mental models. This is an important insight that goes beyond the nature of the data by factoring in the social dimension of how the data is used in the home. Energy-centric (rather than time-centric) representations are in line with how householders think of energy. They do not think of their dishwasher as having a certain power pattern over time, but as a device that consumes energy to wash their dishes. We found area-based representations to be useful for appliance-level feedback. In addition, pictographs proved useful as they provide all information in one element, instead of having to refer to a key which is needed for other graphs, such as bar charts or line graphs. The findings challenge the common practice to show time series data to householders.

In summary, this thesis refines and advances our understanding of how users read and make sense of domestic electricity data. Numerous research studies have previously found that smart infrastructures are not reaching their full potential (both aggregated and disaggregated tools) (Darby, 2006; Kelly & Knottenbelt, 2016). However, field studies measuring changes in consumption typically look at averages of groups. First, they often do not assess and account for the potential of households to save (if a household uses little energy to start with, they have limited possibilities to further decrease consumption and so the study cannot find an effect). Second, the lack of behavioural change observed in previous studies might be due to some extent to the feedback not being smart enough yet and householders not gaining sufficient insights.

In contrast to previous studies, where effects were limited and participants sometimes disengaged, participants in Study 3 were keen to learn and participants in Studies 4 and 6 identified opportunities for change (or confirmed that they were not consuming much). This thesis, and other very recent studies, suggest that maybe we need to revisit how feedback is given rather than giving up on residential eco-feedback (Mogles et al., 2017; Spence et al., 2018). This thesis establishes appliance-wise disaggregation and usable visualisations as a central user requirement. The reliability of these insights has been confirmed by a range of studies and the mixed methods approach. These are very topical findings, seeing that smart infrastructures are being installed worldwide. We side with Strengers' (2011) view that the conflicting findings to date should not 'lead us to conclude that eco-feedback is ineffective—it can and does achieve significant resource reductions (and every bit surely counts)'.

6.3 Limitations

The first limitation in this thesis is the practical constraint to test disaggregated feedback in big field trials and record energy consumption data. There is still no technical solution to disaggregation and most tools that offer appliance-level feedback have not achieved valid results. Testing inaccurate feedback would have been of limited use, so the studies in this thesis refrained from using tools on the market that did not promise to deliver good results. A related constraint was that sometimes, smart energy systems cannot be installed easily in people's homes. For example, difficulties were encountered when trying to install the Loop kit and other sensors in a few test households for Study 3, and some appliances could not have the necessary smart plugs installed for the monitoring in Study 6.

The second limitation of this thesis is that it did not investigate behaviour change. This thesis is motivated by the need to reduce energy consumption to contribute to the mitigation of climate change. The research conducted in this thesis focused on householders' understanding of smart energy feedback. The studies assessed how energy data visualisations can positively impact householders' knowledge and understanding of how they could save energy. It was beyond the scope of this thesis to show if, or how, increased understanding contributes to actual energy savings, i.e. behaviour change. As has been discussed in Study 1, knowledge and behaviour do not have a deterministic relationship, i.e. just because a person knows how to save energy, does not guarantee that they will do it (Ajzen & Fishbein, 1977). However, as pointed out several decades ago by Mettler-Meibom and Wichmann (1982), people cannot change if they do not know how to change. Knowing how much energy practices in the home consume and how this can be influenced through retrofits, programme settings, and behaviour (Study 6), is a necessary precondition to behaviour change (albeit not sufficient).

The third limitation of this thesis is that the sample of participants who took part in the study was not necessarily representative of the general population. As has been pointed out before in the literature, HCI research is predominantly set within western, educated, industrialised, rich, and democratic societies (Sturm et al., 2015). The lab experiments tested mostly highly educated students, and Study 6 consisted of tech-savvy energy experts. However, even though smart home technologies might not be widely adopted yet, they are becoming more and more ubiquitous and Smart Meters are being rolled out to all households across the UK (and other countries worldwide). The hope is that smart energy feedback helps users cut their consumption, which is important because industrialised countries urgently need to reduce their emissions (Paris Agreement, 2015).

6.4 Future Research

There are aspects of smart energy feedback that require further research. These revolve around questions of how better understanding leads to behaviour change, and the role that smart energy feedback plays in future scenarios.

As discussed above, this research is limited to investigating users' understanding of how much energy they consume and how they could save energy. This does not grant that they reduce their consumption. It can be assumed that the need for cleanliness, comfort, and convenience in the home remain strong incentives to not change ones behaviour (Shove, 2003). Understanding energy consumption and wanting to change it are two separate steps. Research suggests that smart feedback must provide much more than just information on use, but rather support for taking action (Mogles et al., 2017; Spence et al., 2018). The technical challenge here is to obtain the relevant data that can be used to provide actionable tips. In the future, this might be achieved through connected solutions in smart homes with their own Internet of Things.

In addition to reducing consumption, customers need incentives to also shift some activities away from peak periods. 'Economy 7' is a differential tariff that offers regular energy prices during the day, and cheaper rates for seven hours during the night. Many customers in the UK have been on a differential tariff for decades, but smart metering infrastructure broadens the possibilities for more sophisticated time-of-use tariffs. In a 'green' future where energy comes predominantly from clean, renewable sources, it might seem less central to incentivise householders to cut their total consumption. For energy providers, the challenge is to provide the maximum power that is used at a time and to deal with the periods of low usage. Storage of renewable energy is difficult and providers will struggle to balance the load on the grid if all customers keep using energy at peak times. Incentivising householders to shift consumption requires both time-centric understanding (when to do or not to do something) and energy-centric understanding (what to do or not to do at a given time). Research and development should seek to find a balance between providers' needs and householders' needs (Marvin, Chappells, & Guys, 1999), so that both parties may benefit from the collected data.

6.5 Conclusion

This thesis refines and advances our understanding of how users read and make sense of domestic electricity data visualisations. So far, a focus has been placed on providing near real-time feedback. This possibility is afforded by smart meters and allows for immediate feedback, rather than delayed feedback from conventional energy bills. However, it is still difficult to learn from instantaneous feedback, because current near real-time information provides a cumulative figure that summarises total consumption. This requires investigative effort from the householder to find out which practices are contributing towards the total consumption. This is also the case for rich data histories, focusing on the display of power usage over time. Instead, appliance-centric feedback is needed for people to learn how much energy household practices consume. This means, energy-centric feedback is needed, instead of time-centric feedback, that tells users how much energy they consume carrying out everyday practices in the home.

Bibliography

Abrahamse, W., Steg, L., Vlek, C., & Rothengatter, T. (2005). A review of intervention studies aimed at household energy conservation. *Journal of Environmental Psychology*, *25*(3), 273–291.

Ackoff, R. (1989). From data to wisdom, *16*(1), 3–9.

Ajzen, I., & Fishbein, M. (1977). Attitude-Behavior Relations: A Theoretical Analysis and Review of Empirical Research, *84*(5).

Alan, A. T., Costanza, E., Ramchurn, S. D., Fischer, J., Rodden, T., & Jennings, N. R. (2016). Tariff Agent: Interacting with a Future Smart Energy System at Home. *ACM Trans. Comput.-Hum. Interact.*, *23*(4), 25:1–25:28.

Álvarez, P., & Vega, P. (2009). Actitudes ambientales y conductas sostenibles. Implicaciones para la educación ambiental. *Revista de Psicodidáctica*, *14*(2).

Anderson, W., & White, V. (2009a). *Exploring consumer preferences for home energy display functionality* (Report to the Energy Saving Trust No. 123).

Anderson, W., & White, V. (2009b). Exploring consumer preferences for home energy display functionality. *Centre for Sustainable Energy, Tech. Rep*.

Armel, K. C., Gupta, A., Shrimali, G., & Albert, A. (2013). Is disaggregation the holy grail of energy efficiency? The case of electricity. *Energy Policy*, *52*(Supplement C), 213–234.

Aronson, J. (1995). A Pragmatic View of Thematic Analysis, *2*(1), 1–3.

Attari, S. Z., DeKay, M. L., Davidson, C. I., & Bruine de Bruin, W. (2010). Public perceptions of energy consumption and savings. *Proceedings of the National Academy of Sciences*, *107*(37), 16054–16059.

Bager, S. (2014). *Information Feedback, Behaviour and 'Smart Meters.'* Lund, Sweden.

Balchin, W. G. V., & Coleman, A. M. (1966). Graphicacy should be the fourth ace in the pack. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *3*(1), 23–28.

Batra, N., Kelly, J., Parson, O., Dutta, H., Knottenbelt, W., Rogers, A., … Srivastava, M. (2014). NILMTK: an open source toolkit for non-intrusive load monitoring. In *Proceedings of the 5th international conference on Future energy systems* (pp. 265–276). ACM.

Baur, D., Lee, B., & Carpendale, S. (2012). TouchWave: Kinetic Multi-touch Manipulation for Hierarchical Stacked Graphs. In *Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces* (pp. 255–264). New York, NY, USA: ACM.

Bertin, J., & Barbut, M. (1973). Sémiologie graphique.

Bird, J., Fozzati, D., Harrison, D., & Marshall, P. (2013). Healthy Shopping: A Longitudinal Study of a Mobile App to Encourage a Balanced Diet. In *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication* (pp. 1047–1054). New York, NY, USA: ACM.

Blandford, A., Furniss, D., & Makri, S. (2016). Qualitative HCI Research: Going Behind the Scenes, *9*(1), 1–115.

Borghouts, J., Soboczenski, F., Cairns, P., & Brumby, D. (2015). Visualizing Magnitude: Graphical Number Representations Help Users Detect Large Number Entry Errors. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *59*(1), 591–595.

Boy, J., Rensink, R. A., Bertini, E., & Fekete, J. D. (2014). A Principled Way of Assessing Visualization Literacy. *IEEE Transactions on Visualization and Computer Graphics*, *20*(12), 1963–1972.

Brewer, R. S. (2013). *Fostering sustained energy behavior change and increasing energy literacy in a student housing energy challenge*. University of Hawai'i at Manoa.

Broms, L., Ehrnberger, K., Ilstedt Hejlm, S., & Bång, M. (2009). The Energy AWARE Clock : Incorporating Electricity Use in the Social Interactions of Everyday Life. In *DIVA*.

Brounen, D., Kok, N., & Quigley, J. (2013). Energy literacy, awareness, and conservation behavior of residential households - ScienceDirect, *38*, 42–50.

Brynjarsdottir, H., H\a akansson, M., Pierce, J., Baumer, E., DiSalvo, C., & Sengers, P. (2012). Sustainably Unpersuaded: How Persuasion Narrows Our Vision of Sustainability. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 947–956). New York, NY, USA: ACM.

Busby, J. S., & Chung, P. W. H. (2003). In What Ways are Designers' and Operators' Reasonable-World Assumptions Not Reasonable Assumptions? *Process Safety and Environmental Protection*, *81*(2), 114–120.

Buxton, B. (2010). *Sketching User Experiences: Getting the Design Right and the Right Design*. Morgan Kaufmann.

Cairns, P., & Cox, A. (2008). *Research methods for human-computer interaction* (Vol. 12). Cambridge: Cambridge University Press.

Camerer, C. F., & Hogarth, R. M. (1999). The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework. *Journal of Risk and Uncertainty*, *19*(1–3), 7–42.

Carpendale, S. (2013). Innovations in Visualization. In *Proceedings of Graphics Interface 2013* (pp. 1–8). Toronto, Ont., Canada, Canada: Canadian Information Processing Society.

Cheng, P. (2011). Probably Good Diagrams for Learning: Representational Epistemic Recodification of Probability Theory, *3*(3), 475–498.

Cheng, P., & Barone, R. (2017). Representing Complex Problems: A Representational Epistemic Approach. In *Learning to solve complex scientific problems* (pp. 97–130). Routledge.

Cheng, P. C.-H. (2014). Graphical notations for syllogisms: How alternative representations impact the accessibility of concepts. *Journal of Visual Languages & Computing*, *25*(3), 170–185.

Cheng, P. C.-H., Lowe, R. K., & Scaife, M. (2001). Cognitive Science Approaches To Understanding Diagrammatic Representations. In *Thinking with Diagrams* (pp. 79–94). Springer, Dordrecht.

Chetty, M., Tran, D., & Grinter, R. E. (2008). Getting to green: understanding resource consumption in the home. In *Proceedings of the 10th international conference on Ubiquitous computing* (pp. 242–251). ACM.

Chiang, T., Mevlevioglu, G., Natarajan, S., Padget, J., & Walker, I. (2014). Inducing [sub]conscious energy behaviour through visually displayed energy information: A case study in university accommodation. *Energy and Buildings*, *70*, 507–515.

Chiang, T., Natarajan, S., & Walker, I. (2012). A laboratory test of the efficacy of energy display interface design. *Energy and Buildings*, *55*(Supplement C), 471–480.

Chisik, Y. (2011). An image of electricity: towards an understanding of how people perceive electricity. *Human-Computer Interaction–INTERACT 2011*, 100–117.

Clarke, V., & Braun, V. (2014). Thematic Analysis. In *Encyclopedia of Quality of Life and Well-Being Research* (pp. 6626–6628). Springer, Dordrecht.

Cleveland, W. S., & McGill, R. (1984). Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *Journal of the American Statistical Association*, *79*(387), 531–554.

Consolvo, S., McDonald, D. W., Toscos, T., Chen, M. Y., Froehlich, J., Harrison, B., … Landay, J. A. (2008). Activity Sensing in the Wild: A Field Trial of Ubifit Garden. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1797–1806). New York, NY, USA: ACM.

Cosar-Jorda, P., Buswell, R. A., & Mitchell, V. (2013). Estimating the potential reductions in energy demand through efficiency, control and lifestyle change in a real home.

Costanza, E., Ramchurn, S. D., & Jennings, N. R. (2012). Understanding Domestic Energy Consumption Through Interactive Visualisation: A Field Study. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (pp. 216–225). New York, NY, USA: ACM.

Cox, A. L., Bird, J., & Fleck, R. (2013). Digital Epiphanies: how self-knowledge can change habits and our attitudes towards them. In *Presented at: The 27th International British Computer Society Human Computer Interaction Conference: The Internet of things, Brunel University, London, UK. (2013)*. Brunel University, London, UK.

Darby, S. (2001). Making it obvious: designing feedback into energy consumption. In *Energy efficiency in household appliances and lighting* (pp. 685–696). Springer.

Darby, S. (2006). The effectiveness of feedback on energy consumption. A review for DEFRA of the literature on metering, billing and direct displays. Environmental Change Institute, University of Oxford. *A Review for DEFRA of the Literature on Metering Billing and Direct Displays, 22nd*.

Darby, S. (2010). Smart metering: what potential for householder engagement? *Building Research & Information*, *38*(5), 442–457.

Davies, N., Clinch, S., Mikusz, M., Bates, O., Turner, H., & Friday, A. (2017). Better off: When Should Pervasive Displays Be Powered Down? In *Proceedings of the 6th ACM International Symposium on Pervasive Displays* (pp. 19:1–19:9). New York, NY, USA: ACM.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science*, *35*(8), 982–1003.

DECC (2009). Impact Assessment of a GB-wide Smart Meter Roll Out for the Domestic Sector. DECC, London.

DECC (2013). Smart Metering Implementation Programme. Smart Metering Equipment Technical Specifications Version 2. DECC, London.

DECC (2015). Smart Metering Early Learning Project: Consumer Survey and qualitative research. DECC, London.

Delmas, M. A., Fischlein, M., & Asensio, O. I. (2013). Information strategies and energy conservation behavior: A meta-analysis of experimental studies from 1975 to 2012. Energy Policy, 61, 729–739.

DeWaters, J. E., & Powers, S. E. (2011). Energy literacy of secondary students in New York State (USA): A measure of knowledge, affect, and behavior. *Energy Policy*, *39*(3), 1699–1710.

DeWaters, J., Powers, S., & Graham, M. (2007). Developing an energy literacy scale.

DeWaters, J., Qaqish, B., Graham, M., & Powers, S. (2013). Designing an Energy Literacy Questionnaire for Middle and High School Youth. *The Journal of Environmental Education*, *44*(1), 56–78.

Dickson, D. (2005). The case for a “deficit model” of science communication, *27*.

Dütschke, E., Unterländer, M., & Wietschel, M. (2012). *Variable Stromtarife aus Kundensicht: Akzeptanzstudie auf Basis einer Conjoint-Analyse* (Working Paper No. S1/2012). Working paper sustainability and innovation.

Ehrhardt-Martinez, K., Donnelly, K., & Laitner, S. (2010). *Advanced Metering Initiatives and Residential Feedback Programs: A Meta-Review for Household Electricity-Saving Opportunities* (p. 140). Washington, DC: American Council for an Energy-Efficient Economy.

Entwistle, J. M., Rasmussen, M. K., Verdezoto, N., Brewer, R. S., & Andersen, M. S. (2015). Beyond the Individual: The Contextual Wheel of Practice As a Research Framework for Sustainable HCI. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 1125–1134). New York, NY, USA: ACM.

Epstein, D. A., Ping, A., Fogarty, J., & Munson, S. A. (2015). A Lived Informatics Model of Personal Informatics. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 731–742). New York, NY, USA: ACM.

Fischer, C. (2008). Feedback on household electricity consumption: a tool for saving energy? *Energy Efficiency*, *1*(1), 79–104.

Fischer, J. E., Ramchurn, S. D., Osborne, M., Parson, O., Huynh, T. D., Alam, M., … Jennings, N. R. (2013). Recommending Energy Tariffs and Load Shifting Based on Smart Household Usage Profiling. In *Proceedings of the 2013 International Conference on Intelligent User Interfaces* (pp. 383–394). New York, NY, USA: ACM. h

Fleck, R., & Fitzpatrick, G. (2010). Reflecting on Reflection: Framing a Design Landscape. In *Proceedings of the 22Nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction* (pp. 216–223). New York, NY, USA: ACM.

Fogg, B. J. (2002). Persuasive Technology: Using Computers to Change What We Think and Do. *Ubiquity*, *2002*(December). https://doi.org/10.1145/764008.763957

Ford, R., & Karlin, B. (2013). Graphical Displays in Eco-Feedback: A Cognitive Approach. In *Design, User Experience, and Usability. Web, Mobile, and Product Design* (pp. 486–495). Springer, Berlin, Heidelberg.

Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making Sense of Graphs: Critical Factors Influencing Comprehension and Instructional Implications. *Journal for Research in Mathematics Education*, *32*(2), 124–158.

Froehlich, J. (2011). *Sensing and feedback of everyday activities to promote environmental behaviors*. University of Washington.

Froehlich, J. E., Kay, M., Larsen, J. E., & Thomaz, E. (2014). Disasters in Personal Informatics: The Unpublished Stories of Failure and Lessons Learned. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (pp. 673–678). New York, NY, USA: ACM.

Froehlich, J., Larson, E., Gupta, S., Cohn, G., Reynolds, M., & Patel, S. (2011). Disaggregated End-Use Energy Sensing for the Smart Grid. *IEEE Pervasive Computing*, *10*(1), 28–39.

Fry, E. (1981). Graphical Literacy. *Journal of Reading*, *24*(5), 383–389.

Gabe-Thomas, E., Walker, I., Verplanken, B., & Shaddick, G. (2016). Householders' Mental Models of Domestic Energy Consumption: Using a Sort-And-Cluster Method to Identify Shared Concepts of Appliance Similarity, *11*(7).

Galesic, M., & Garcia-Retamero, R. (2011). Graph Literacy: A Cross-Cultural Comparison. *Medical Decision Making*, *31*(3), 444–457.

Goncalves, H., Ocneanu, A., Berges, M., & Fan, R. H. (2011). Unsupervised disaggregation of appliances using aggregated consumption data. In *The 1st KDD Workshop on Data Mining Applications in Sustainability (SustKDD)*.

Gonzalez, C., Thomas, R. P., & Vanyukov, P. (2005). The relationships between cognitive ability and dynamic decision making. *Intelligence*, *33*(2), 169–186.

Goulden, M., Bedwell, B., Rennick-Egglestone, S., Rodden, T., & Spence, A. (2014). Smart grids, smart users? The role of the user in demand side management. *Energy Research & Social Science*, *2*, 21–29.

Greenberg, S., Carpendale, S., Marquardt, N., & Buxton, B. (2011). *Sketching User Experiences: The Workbook*. Elsevier.

Greis, M., Henze, N., & Schmidt, A. (2015). Predicting the Future: Towards Personal Simulation. In *CHI*. Seoul, South Korea.

Hargreaves, T. (2011). Practice-ing behaviour change: Applying social practice theory to pro-environmental behaviour change.

Hargreaves, T. (2018). Beyond energy feedback. *Building Research & Information*, *46*(3), 332–342.

Hargreaves, T., Nye, M., & Burgess, J. (2010). Making energy visible: A qualitative field study of how householders interact with feedback from smart energy monitors. *Energy Policy*, *38*(10), 6111–6119.

Haroz, S., Kosara, R., & Franconeri, S. L. (2015). ISOTYPE Visualization: Working Memory, Performance, and Engagement with Pictographs. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 1191–1200). New York, NY, USA: ACM.

Harrison, D., Bird, J., Marshall, P., & Berthouze, N. (2013). Looking for bright spots: a bottom-up approach to encouraging urban exercise. *BCS (The Chartered Institute for IT).*

Harrison, D., Marshall, P., Bianchi-Berthouze, N., & Bird, J. (2015). Activity Tracking: Barriers, Workarounds and Customisation. In *Proceedings of the 2015 ACM*

International Joint Conference on Pervasive and Ubiquitous Computing (pp. 617–621). New York, NY, USA: ACM.
Hart, G. W. Natick, Kern, J., & Schweppe, F. C. Carlisle. (1989). *Non-Intrusive Appliance Monitor Apparatus* (No. US 4858141). MASSACHUSETTS INST OF TECH.
Hekler, E. B., Klasnja, P., Froehlich, J. E., & Buman, M. P. (2013). Mind the Theoretical Gap: Interpreting, Using, and Developing Behavioral Theory in HCI Research. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3307–3316). New York, NY, USA: ACM.
Hofman, H. (1980). Energy Crisis—Schools To The Rescue Again. *School Science and Mathematics*, *80*(6), 467–478.
Holtzblatt, K., & Jones, S. (1993). Contextual inquiry: A participatory technique for system design. In *Participatory design: Principles and practices* (pp. 177–210).
Hungerford, H. R., & Volk, T. L. (1990). Changing Learner Behavior Through Environmental Education. *The Journal of Environmental Education*, *21*(3), 8–21.
Katzeff, C., Wessman, S., & Colombo, S. (2017). "Mama, It's Peacetime!": Planning, Shifting, and Designing Activities in the Smart Grid Scenario. *Proceedings of the Conference on Design and Semantics of Form and Movement - Sense and Sensitivity, DeSForM 2017*.
Katzev, R. D., & Johnson, T. R. (1987). *Promoting energy conservation: An analysis of behavioral research*. United States.
Kay, M., Choe, E. K., Shepherd, J., Greenstein, B., Watson, N., Consolvo, S., & Kientz, J. A. (2012). Lullaby: A Capture & Access System for Understanding the Sleep Environment. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (pp. 226–234). New York, NY, USA: ACM.
Kelly, J., & Knottenbelt, W. (2015). The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes. *Scientific Data*, *2*, 150007.
Kelly, J., & Knottenbelt, W. (2016). Does disaggregated electricity feedback reduce domestic electricity consumption? A systematic review of the literature. *ArXiv:1605.00962 [Cs]*.
Kempton, W., & Layne, L. L. (1994). The consumer's energy analysis environment. *Energy Policy*, *22*(10), 857–866.
Kempton, W., & Montgomery, L. (1982). Folk quantification of energy. *Energy*, *7*(10), 817–827.
Kidd, A., & Williams, P. (2008). The Talybont trial: exploring the psychology of smart meters. Brecon, Wales, The Prospectory.
Kirsh, D. (2010). Thinking with external representations, *25*(4), 441–454.
Knowles, B., Blair, L., Coulton, P., & Lochrie, M. (2014). Rethinking Plan A for Sustainable HCI. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems* (pp. 3593–3596). New York, NY, USA: ACM.
Krishnamurti, T., Davis, A. L., Wong-Parodi, G., Wang, J., & Canfield, C. (2013). Creating an in-home display: Experimental evidence and guidelines for design. *Applied Energy*, *108*, 448–458.
Krueger, R., & Casey, M. A. (2014). *Focus groups: A practical guide for applied research*. Sage publications.

Landsberger, H. A. (1958). Hawthorne Revisited: Management and the Worker, Its Critics, and Developments in Human Relations in Industry.

Larkin, J. H., & Simon, H. A. (1987). Why a Diagram is (Sometimes) Worth Ten Thousand Words. *Cognitive Science*, *11*(1), 65–100.

Latham, G., & Locke, E. (1991). Self-regulation through goal setting - ScienceDirect, *50*(2), 212–247.

Lazar, A., Koehler, C., Tanenbaum, J., & Nguyen, D. H. (2015). Why We Use and Abandon Smart Devices. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 635–646). New York, NY, USA: ACM.

Lewis, C., & Rieman, J. (1993). *Task-centered user interface design. A practical introduction*.

Leygue, C., Ferguson, E., Skatova, A., & Spence, A. (2014). Energy Sharing and Energy Feedback: Affective and Behavioral Reactions to Communal Energy Displays. *Frontiers in Energy Research*, *2*.

Li, I., Dey, A., & Forlizzi, J. (2010). A Stage-based Model of Personal Informatics Systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 557–566). New York, NY, USA: ACM.

Lin, J. J., Mamykina, L., Lindtner, S., Delajoux, G., & Strub, H. B. (2006). Fish'n'Steps: Encouraging Physical Activity with an Interactive Computer Game. In *UbiComp 2006: Ubiquitous Computing* (pp. 261–278). Springer, Berlin, Heidelberg.

Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General Performance on a Numeracy Scale among Highly Educated Samples , General Performance on a Numeracy Scale among Highly Educated Samples. *Medical Decision Making*, *21*(1), 37–44.

Locoro, A., Cabitza, F., Actis-Grosso, R., & Batini, C. (2017). Static and interactive infographics in daily tasks: A value-in-use and quality of interaction user study. *Computers in Human Behavior*, *71*, 240–257.

Loorak, M. H., Perin, C., Kamal, N., Hill, M., & Carpendale, S. (2016). TimeSpan: Using Visualization to Explore Temporal Multi-dimensional Data of Stroke Patients. *IEEE Transactions on Visualization and Computer Graphics*, *22*(1), 409–418.

Lovett, T., Gabe-Thomas, E., Natarajan, S., O'Neill, E., & Padget, J. (2013). "Just Enough" Sensing to ENLITEN: A Preliminary Demonstration of Sensing Strategy for the "Energy Literacy Through an Intelligent Home Energy Advisor" (ENLITEN) Project. In *Proceedings of the Fourth International Conference on Future Energy Systems* (pp. 279–280). New York, NY, USA: ACM.

Mamykina, L., Mynatt, E., Davidson, P., & Greenblatt, D. (2008). MAHI: Investigation of Social Scaffolding for Reflective Thinking in Diabetes Management. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 477–486). New York, NY, USA: ACM.

Maréchal, K., & Holzemer, L. (2015). Getting a (sustainable) grip on energy consumption: The importance of household dynamics and 'habitual practices.' *Energy Research & Social Science*, *10*, 228–239.

Marvin, S., Chappells, H., & Guys, S. (1999). Pathways of smart metering development: shaping environmental innovation - ScienceDirect, *23*(2), 109–126.

McCalley, L. T., & Midden, C. J. H. (2002). Energy conservation through product-integrated feedback: The roles of goal-setting and social orientation. *Journal of Economic Psychology*, *23*(5), 589–603.

Mettler-Meibom, B., & Wichmann, B. (1982). Informationsstand und Einstellung als Verhaltensregulative. In *Einfluß des Verbraucherverhaltens auf den Energiebedarf privater Haushalte* (pp. 73–88). Springer, Berlin, Heidelberg.

Mogles, N., Walker, I., Ramallo-González, A. P., Lee, J., Natarajan, S., Padget, J., … Coley, D. (2017). How smart do smart meters need to be? *Building and Environment*, *125*(Supplement C), 439–450.

Morgan, D. (1997). *The focus group guidebook* (Vol. 1). Sage publications.

Munzner, T. (2014). *Visualization analysis and design*. AK Peters/CRC Press.

Murugesan, L. K., Hoda, R., & Salcic, Z. (2014). Visualization of Electricity Consumption: Software Prototype through Literature Survey, *3*(3).

Murugesan, L. K., Hoda, R., & Salcic, Z. (2015). Design criteria for visualization of energy consumption: A systematic literature review. *Sustainable Cities and Society*, *18*, 1–12.

Neustaedter, C., Bartram, L., & Mah, A. (2013). Everyday Activities and Energy Consumption: How Families Understand the Relationship. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1183–1192). New York, NY, USA: ACM.

Osman, M. (2011). *Controlling Uncertainty: Decision Making and Learning in Complex Worlds*. John Wiley & Sons.

Palm, J., & Ellegård, K. (2011). Visualizing energy consumption activities as a tool for developing effective policy. *International Journal of Consumer Studies*, *35*(2), 171–179.

Papert, S. (1980). *Mindstorms: Children, Computers, and Powerful Ideas*. New York, NY, USA: Basic Books, Inc.

Paris Agreement. (2015). United Nations.

Patel, S. N., Robertson, T., Kientz, J. A., Reynolds, M. S., & Abowd, G. D. (2007). At the flick of a switch: Detecting and classifying unique electrical events on the residential power line. *Lecture Notes in Computer Science*, *4717*, 271–288.

Peebles, D., Ramduny-Ellis, D., Ellis, G., & Bonner, J. V. (2013). The influence of graph schemas on the interpretation of unfamiliar diagrams.

Pereira, L., Quintal, F., Barreto, M., & Nunes, N. J. (2013). Understanding the Limitations of Eco-feedback: A One-Year Long-Term Study. In *Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data* (pp. 237–255). Springer, Berlin, Heidelberg.

Pierce, J., Fan, C., Lomas, D., Marcu, G., & Paulos, E. (2010). Some Consideration on the (in)Effectiveness of Residential Energy Feedback Systems. In *Proceedings of the 8th ACM Conference on Designing Interactive Systems* (pp. 244–247). New York, NY, USA: ACM.

Pinker, S. (1990). A theory of graph comprehension. *Artificial Intelligence and the Future of Testing*, 73–126.

Ploderer, B., Reitberger, W., Oinas-Kukkonen, H., & Gemert-Pijnen, J. van. (2014). Social interaction and reflection for behaviour change. *Personal and Ubiquitous Computing*, *18*(7), 1667–1676.

Pousman, Z., Stasko, J., & Mateas, M. (2007). Casual Information Visualization: Depictions of Data in Everyday Life, *13*(6), 1145–1152.

Prochaska, J., & DiClemente, C. (2005). The transtheoretical approach. In *Handbook of psychotherapy integration* (2nd ed., pp. 147–171).

Prost, S., Mattheiss, E., & Tscheligi, M. (2015). From Awareness to Empowerment: Using Design Fiction to Explore Paths Towards a Sustainable Energy Future. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 1649–1658). New York, NY, USA: ACM.

Purpura, S., Schwanda, V., Williams, K., Stubler, W., & Sengers, P. (2011). Fit4Life: The Design of a Persuasive Technology Promoting Healthy Behavior and Ideal Weight. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 423–432). New York, NY, USA: ACM.

Rabiee, F. (2004). Focus-group interview and data analysis. *Proceedings of the nutrition society*, 63(4), 655-660.

Rasmussen, M. K. (2016). People, Technology, and Energy: Configurations of Practices and Place. In *Smart for Life.*

Rego Teixeira, A. (2014). *LE « SMART-METERING » EN SUISSE ROMANDE : conditions sociales d'acceptation et d'appropriation*. University of Geneva. Retrieved from

Reinhardt, A., Baumann, P., Burgstahler, D., Hollick, M., Chonov, H., Werner, M., & Steinmetz, R. (2012). On the accuracy of appliance identification based on distributed load metering data. In *Sustainable Internet and ICT for Sustainability (SustainIT), 2012* (pp. 1–9). IEEE.

Ritchie, J., Lewis, J., McNaughton Nichols, C., & Ormston, R. (2014). *Qualitative Research Practice. A Guide for Social Science Students & Researchers* (2nd ed.). London, UK.: Sage publications.

Ritchie, J. & Spencer, L. (1994). Qualitative data analysis for applied policy research. In *Analysing Qualitative Data*, pp. 173-194. London: Routledge.

Roberts, S., & Baker, W. (2003). Towards effective energy information. *Improving Consumer Feedback on Energy Consumption, Centre for Sustainable Energy, Bristol, UK*.

Rogers, Y., Yuill, N., & Marshall, P. (2013). Contrasting lab- based and in-the-wild studies for evaluating multi-user technologies. In *The Sage Handbook of Digital Technology Research*. London, UK.

Rooksby, J., Rost, M., Morrison, A., & Chalmers, M. (2014). Personal Tracking As Lived Informatics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1163–1172). New York, NY, USA: ACM.

Schlager, M. (2015). Curating Family Life in the Smart Home, 4.

Schön, D. A. (1987). *Educating the reflective practitioner*. San Francisco: Jossey-Bass Publishers, 350 Sansome Street, San Francisco, CA 94104 ($24.

Schrader, P. G., & Lawless, K. A. (2004). The Knowledge, Attitudes, & Behaviors Approach: How to Evaluate Performance and Learning in Complex Environments. *Performance Improvement*, *43*(9), 8–15.

Schwartz, T., Denef, S., Stevens, G., Ramirez, L., & Wulf, V. (2013). Cultivating Energy Literacy: Results from a Longitudinal Living Lab Study of a Home Energy Management System. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1193–1202). New York, NY, USA: ACM.

Schwartz, T., Stevens, G., Jakobi, T., Denef, S., Ramirez, L., Wulf, V., & Randall, D. (2015). What People Do with Consumption Feedback: A Long-Term Living Lab Study of a Home Energy Management System. *Interacting with Computers*, *27*(6), 551–576.

Schwartz, T., Stevens, G., Ramirez, L., & Wulf, V. (2013). Uncovering Practices of Making Energy Consumption Accountable: A Phenomenological Inquiry. *ACM Trans. Comput.-Hum. Interact.*, *20*(2), 12:1–12:30.

Sedlmair, M., Meyer, M., & Munzner, T. (2012). Design Study Methodology: Reflections from the Trenches and the Stacks. *IEEE Transactions on Visualization & Computer Graphics*, *18*(12), 2431–2440.

Seidman, I. (2013). *Interviewing as Qualitative Research: A Guide for Researchers in Education and the Social Sciences*. Teachers College Press.

Shove, E. (2003). *Comfort, Cleanliness and Convenience: The Social Organization of Normality*. Bloomsbury Academic.

Sokoloski, R. (2015). *Disaggregated Electricity Consumption: Using Appliance-Specific Feedback to Promote Energy Conservation* (Thesis).

Sorrell, S., Dimitropoulos, J., & Sommerville, M. (2009). Empirical estimates of the direct rebound effect: A review. *Energy Policy*, *37*(4), 1356–1371.

Spence, A., Leygue, C., Bedwell, B., & O'Malley, C. (2014). Engaging with energy reduction: Does a climate change frame have the potential for achieving broader sustainable behaviour? *Journal of Environmental Psychology*, *38*, 17–28.

Spence, A., Goulden, M., Leygue, C., Banks, N., Bedwell, B., Jewell, M., Yang, R. & Ferguson, E. (2018). Digital energy visualizations in the workplace: the e-Genie tool. *Building Research & Information*, *46*(3), 272–283.

Strengers, Y. (2011a). Designing eco-feedback systems for everyday life. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2135–2144). ACM.

Strengers, Y. (2011b). Negotiating everyday life: The role of energy and water consumption feedback , Negotiating everyday life: The role of energy and water consumption feedback. *Journal of Consumer Culture*, *11*(3), 319–338.

Strengers, Y. (2014). Smart Energy in Everyday Life: Are You Designing for Resource Man? *Interactions*, *21*(4), 24–31.

Sturm, C., Oh, A., Linxen, S., Abdelnour Nocera, J., Dray, S., & Reinecke, K. (2015). How WEIRD is HCI?: Extending HCI Principles to Other Countries and Cultures. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 2425–2428). New York, NY, USA: ACM.

Sweeney, J. C., Kresling, J., Webb, D., Soutar, G. N., & Mazzarol, T. (2013). Energy saving behaviours: Development of a practice-based model. *Energy Policy*, *61*, 371–381.

Tong, X., Gromala, D., Bartram, L., Rajabiyazdi, F., & Carpendale, S. (2015). Evaluating the effectiveness of three physical activity visualizations—how people perform vs. perceive. In *IEEE VIS*.

Tufte, E. R. (1983). The visual display of quantitative infor- mation. Connecticut, CT: Graphics Press.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, *185*(4157), 1124–1131.

Tversky, B., Corter, J. E., Nickerson, J. V., Zahner, D., & Rho, Y. J. (2008). Transforming Descriptions and Diagrams to Sketches in Information System Design. In *Diagrammatic Representation and Inference* (pp. 242–256). Springer, Berlin, Heidelberg.

Van Dam, S. s., Bakker, C. a., & Van Hal, J. d. m. (2012). Insights into the design, use and implementation of home energy management systems. *Journal of Design Research*, *10*(1–2), 86–101.

Vanhulst, P., & Lalanne, D. (2015). Collaborative Visualizations of Self-impacted Data (p. 4). Presented at the IEEE VIS.

Vaughn, S., Schumm, J., & Sinagub, J. (1996). *Focus group interviews in education and psychology*. Sage publications.

Wallenborn, G., Orsini, M., & Vanhaverbeke, J. (2011). Household appropriation of electricity monitors. *International Journal of Consumer Studies*, *35*, 146–152.

Walny, J., Huron, S., & Carpendale, S. (2015). An Exploratory Study of Data Sketching for Visual Representation. *Computer Graphics Forum*, *34*(3), 231–240.

Winner, L. (1980). Do Artifacts Have Politics? *Daedalus*, *109*(1), 121–136.

Wolff, A., Gooch, D., Cavero Montaner, J. J., Rashid, U., & Kortuem, G. (2016). Creating an Understanding of Data Literacy for a Data-driven Society. *Journal of Community Informatics*, *12*, 9–26.

Wood, G., & Newborough, M. (2007). Energy-use information transfer for intelligent homes: Enabling energy conservation with central and local displays. *Energy and Buildings*, *39*(4), 495–503.

Yang, R., Newman, M. W., & Forlizzi, J. (2014). Making sustainability sustainable: challenges in the design of eco-interaction technologies (pp. 823–832). ACM Press.

Yang, S., Shipworth, M., & Huebner, G. (2015). His, hers or both's? The role of male and female's attitudes in explaining their home energy use behaviours. *Energy and Buildings*, *96*, 140–148.

Yun, T.-J., Jeong, H. Y., Kinnaird, P., Choi, S., Kang, N., & Abowd, G. D. (2010). Domestic Energy Displays: An Empirical Investigation.

Zeifman, M., & Roth, K. (2011). Nonintrusive appliance load monitoring: Review and outlook. *IEEE Transactions on Consumer Electronics*, *57*(1), 76–84.

Zhang, J., & Norman, D. (1994). Representations in Distributed Cognitive Tasks, *18*(1), 87–122.

Zhao, C., Froehlich, J., & Landay, J. A. (2010). Exploring the Cross-Cultural Effects of Feedback for Pro-Environmental Behavior Change (Vol. 14). Redmond, Washington: Microsoft Environmental Research Workshop.

Appendices

Appendix A

The following definitions of energy literacy are exact transcriptions of participants' notes from Activity I in the focus group (Study 1). The academics' group is one note short of the complete list (as one participant took notes on their piece of paper and took the sheet with them at the end of the session).

A1: Understanding of how energy is used and where it comes from. Awareness of energy issues e.g. climate change, need to reduce CO_2 emissions. Practical understanding e.g. of major energy uses, what determines how much energy [is] used to heat home etc.

A2: Understanding the conflict of using and conserving (and wasting) electricity and gas. Could mean understanding of kW & kWh. Understanding concepts of generation (where energy comes from, how it is produced), transmission, challenges of storage.

A3: Understanding one's own energy consumption – how much used over, say 12 months: Understanding where consumed energy comes from. Understanding what appliances use what types of energy.

A4: Knowing 1) how energy is consumed 2) how energy is produced 3) the consequences of the previous two.

A5: Understanding about energy production & consumption. Understand what energy means? Ability to describe energy use in one's daily activities.

A6: What is energy literacy. I come with zero knowledge of the literacy, so: The degree of technical competency relating to the provision of energy services. The tools at a user's disposal to alter the conditions of their access to and interaction with energy services. Knowing how energy services work.

U1: Being 'be-red' [literate, well-read from German 'belesen'] about energy consumption. Knowing how much energy is used by some process or activity. Knowing about what contributes to energy usage.

U2: Understanding energy (sources) options, about how it is transmitted from 'suppliers' to customers, about its use worldwide and nationally, and one's consumption/use of it in everyday life and its effect financially, socially, environmentally…

U3: Understanding and applying uses of energy that consider the levels of energy consumption. Perhaps also use with the goal of 'least consumption' and sustainable use.

U4: Understanding how, when and where energy is used. The common measurements for energy (Kwtt) (miles per gallon). Being able to understand your energy bills and make changes to behaviour to reduce energy consumption.

U5: Ability to 1) + 2). 1) What is energy? -> What kind of types exist? 2) Literacy? -> read – theoretical knowledge -> understand, reflect, decode – practical knowledge -> act, change behaviour.

U6: Ability to understand energy consumption. Also about controlling it? Transfer skills from one situation to another?

U7: The extent to which you can interpret/make sense of energy usage data. Can express your energy usage. Are aware of translating energy usage data and sources (appliances, actions, etc.).

I1: The terminology/language/words that is used to describe 'Energy'.

I2: Literacy – understanding… meaning, context, implications… of my energy usage, consumption, habits.

I3: Understanding how consumer uses energy. Interpretation into type of consumers or load at meter point.

I4: Understanding of the energy I use: where it comes from, what I use it for, how much I use.

I5: One's understanding of energy and its uses: in their household, in the wider world, in dealing with providers.

I6: Considering the impact that energy has in day to day life. Reliance. Usefulness. Considering how it's generated and the impact of that. Considering how you and others use it: when, how much, why.